



```
QY 61 ACTTCCTGTTTCGATTAGTTAGTGGCAATTTAGTAGAGAAATGACAGACCCCGCATTCT 120
Db 61 ACTTCCTGTTTCGATTAGTTAGTGGCAATTTAGTAGAGAAATGACAGACCCCGCATTCT 120
QY 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
Db 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
QY 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAATGCAATCATCTCTCTGAATATGA 240
Db 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAATGCAATCATCTCTCTGAATATGA 240
QY 241 CATCAACCTGTGATATGATATCAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCCG 300
Db 241 CATCAACCTGTGATATGATATCAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCCG 300
QY 301 TGAAGTCCGAGAGCGGGAATGTTGCGTTGATCTGACTGAGTGGCGGTGAACAAGAGT 360
Db 301 TGAAGTCCGAGAGCGGGAATGTTGCGTTGATCTGACTGAGTGGCGGTGAACAAGAGT 360
QY 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATCAATCAACAAACGGTTCAACC 420
Db 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATCAATCAACAAACGGTTCAACC 420
QY 421 GCGTGAACCTGACGATTCGTGACAGCAACCTTACTGTCCTGATCAATGATCTGGTACTGT 480
Db 421 GCGTGAACCTGACGATTCGTGACAGCAACCTTACTGTCCTGATCAATGATCTGGTACTGT 480
QY 481 CAGCGAAGACGTGCTAGACGTTGAAAGCTTCAAGTTCAATGGTTGGAACTGACATGAA 540
Db 481 CAGCGAAGACGTGCTAGACGTTGAAAGCTTCAAGTTCAATGGTTGGAACTGACATGAA 540
QY 541 CAGCGGTTCTGTATCGGCTCTGATGCGGAGACGATCAAGCTGCGCGAGCGAGTTCCG 600
Db 541 CAGCGGTTCTGTATCGGCTCTGATGCGGAGACGATCAAGCTGCGCGAGCGAGTTCCG 600
QY 601 CGCCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGTGAA 660
Db 601 CGCCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCCGTGTGAATCTGTGAA 660
QY 661 GAAATGACCGGCGCTGAGCTGAAACCGACGACCGTACTGTGAGATGATCCGCG 720
Db 661 GAAATGACCGGCGCTGAGCTGAAACCGACGACCGTACTGTGAGATGATCCGCG 720
QY 721 TATTCGTAACATTTGCAATCTACGCGGATAGCGGAAATCATGCGCAATTCACGG 780
Db 721 TATTCGTAACATTTGCAATCTACGCGGATAGCGGAAATCATGCGCAATTCACGG 780
QY 781 TGAAGGTATCGCTTCTGCGGTGATGGAAGATTAATCGGCTTACCAACCGTCAAAA 840
Db 781 TGAAGGTATCGCTTCTGCGGTGATGGAAGATTAATCGGCTTACCAACCGTCAAAA 840
QY 841 AACGGCGCTTTTGAAGCGGTTTATTTTCAACCTTATTTGCAATACGTAATCTATC 900
Db 841 AACGGCGCTTTTGAAGCGGTTTATTTTCAACCTTATTTGCAATACGTAATCTATC 900
QY 901 GTCCGTTGAACTTCTTTTACGCTTT 927
Db 901 GTCCGTTGAACTTCTTTTACGCTTT 927
```

RESULT 2  
U00096\_46/c  
WPCOMMENT

Sequence split into 47 fragments LOCUS U00096 Accession U00096  
Fragment Name Begin End  
U00096\_00 1 110000  
U00096\_01 100001 210000  
U00096\_02 200001 310000  
U00096\_03 300001 410000  
U00096\_04 400001 510000  
U00096\_05 500001 610000

```
U00096_06 600001 710000
U00096_07 700001 810000
U00096_08 800001 910000
U00096_09 900001 1010000
U00096_10 1000001 1110000
U00096_11 1100001 1210000
U00096_12 1200001 1310000
U00096_13 1300001 1410000
U00096_14 1400001 1510000
U00096_15 1500001 1610000
U00096_16 1600001 1710000
U00096_17 1700001 1810000
U00096_18 1800001 1910000
U00096_19 1900001 2010000
U00096_20 2000001 2110000
U00096_21 2100001 2210000
U00096_22 2200001 2310000
U00096_23 2300001 2410000
U00096_24 2400001 2510000
U00096_25 2500001 2610000
U00096_26 2600001 2710000
U00096_27 2700001 2810000
U00096_28 2800001 2910000
U00096_29 2900001 3010000
U00096_30 3000001 3110000
U00096_31 3100001 3210000
U00096_32 3200001 3310000
U00096_33 3300001 3410000
U00096_34 3400001 3510000
U00096_35 3500001 3610000
U00096_36 3600001 3710000
U00096_37 3700001 3810000
U00096_38 3800001 3910000
U00096_39 3900001 4010000
U00096_40 4000001 4110000
U00096_41 4100001 4210000
U00096_42 4200001 4310000
U00096_43 4300001 4410000
U00096_44 4400001 4510000
U00096_45 4500001 4610000
U00096_46 4600001 463675

Continuation (47 of 47) of U00096 from base 4600001 (U00096 Escherichia coli K-12 MG1655)

Query Match 100.0%; Score 927; DB 1; Length 39675;
Best Local Similarity 100.0%; Pred. No. 2,3e-264;
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCATGTTACGCCGATCATGTTAATTTGACGATGATCAGGACAGTCAAGGACTTTTGT 60
Db 38429 GTCATGTTACGCCGATCATGTTAATTTGACGATGATCAGGACAGTCAAGGACTTTTGT 38370
QY 61 ACTTCGTTTGCATTTAGTTAGCAATTTTGTAGCAAAATGACAGACCCGCAATTCT 120
Db 38369 ACTTCGTTTGCATTTAGTTAGTTAGCAATTTTGTAGCAAAATGACAGACCCGCAATTCT 38310
QY 121 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
Db 38309 TATCGTTGAAGAGAGTTGGTAAACAGCAACAGTTGAAAGTATTTTGAAGCGGAAG 38250
QY 181 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAATGCAATCATCTCTGAATATGA 240
Db 38249 CTATGATGTTTTGGAAGCGACAGATGGCCGGAAATGCAATCATCTCTGAATATGA 38190
QY 241 CATCAACCTGTGATATGATATCAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCGG 300
Db 38189 CATCAACCTGTGATATGATATCAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCGG 38130
QY 301 TGAATCGCGGAGAGAGGGAATGTTGCGTTGATGTTCTGATCTGCGGTGACACAGAGT 360
Db 38129 TGAATCGCGGAGAGAGGGAATGTTGCGTTGATGTTCTGATCTGCGGTGACACAGAGT 38070
QY 361 CGATTAATTTCTCGGCTCTGAAATCGGTGAGATGATCAATCAACAAACGGTTCAACC 420
```

Db 38069 CCAATAAATTCCTGCGCTCGAATTCGGTGCAGATGACTACATCAACCAACCTTCAACC 38010  
Qy 421 GCGTGAATCGACGATTCGTGCACGCAACCTACTGTCCCGTACCATGATCTGGGTACTGT 480  
Db 38009 GCGTGAATCGACGATTCGTGCACGCAACCTACTGTCCCGTACCATGATCTGGGTACTGT 37950  
Qy 481 CAGCGAAGAACGTCGTAAGCTTGAAGCTTCAAGTTCAATGTTGGAACTGGACATCA 540  
Db 37949 CAGCGAAGAACGTCGTAAGCTTGAAGCTTCAAGTTCAATGTTGGAACTGGACATCA 37890  
Qy 541 CAGCGCTGTGATGAGCGCCCTGATGCGGACGATCAAGTCGCGCGACGAGATTCCG 600  
Db 37889 CAGCGCTGTGATGAGCGCCCTGATGCGGACGATCAAGTCGCGCGACGAGATTCCG 37830  
Qy 601 CCGCATGCTTCACTTCTGTGAAAACCGAGCAAAATTCAGTCCCGTCTGAATCTGTGAA 660  
Db 37829 CCGCATGCTTCACTTCTGTGAAAACCGAGCAAAATTCAGTCCCGTCTGAATCTGTGAA 37770  
Qy 661 GAAATGACCGCGCGGTGAGCTGAAACCGACGACCGTACTGTAGAGCTGACGATCCGCCG 720  
Db 37769 GAAATGACCGCGCGGTGAGCTGAAACCGACGACCGTACTGTAGAGCTGACGATCCGCCG 37710  
Qy 721 TATTCGTAACATTTGGAATCTACGCGGATAGCGCGGAAATGATGCGCAACATTCACCG 780  
Db 37709 TATTCGTAACATTTGGAATCTACGCGGATAGCGCGGAAATGATGCGCAACATTCACCG 37650  
Qy 781 TGAAGGTTATCGCTTCTGCGGTGATCGAAGATTAATCGCTTACACCGTCAAAAA 840  
Db 37649 TGAAGGTTATCGCTTCTGCGGTGATCGAAGATTAATCGCTTACACCGTCAAAAA 37590  
Qy 841 AACGGCGCTTTTAAAGCGCGCTTTTATTTTCAACCTTATTCAGATACGTAATCTGATC 900  
Db 37589 AACGGCGCTTTTAAAGCGCGCTTTTATTTTCAACCTTATTCAGATACGTAATCTGATC 37530  
Qy 901 GTCCGTTGTAATCTTCTTACTGCGCTT 927  
Db 37529 GTCCGTTGTAATCTTCTTACTGCGCTT 37503

RESULT 3  
ECOW93/c 338534 bp DNA linear BCT 30-JAN-2001  
DEFINITION Escherichia coli K-12 chromosomal region from 92.8 to 00.1 minutes.  
ACCESSION U14003  
VERSION U14003.1 GI:1263172  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
AUTHORS 1 (bases 1 to 338534)  
Blattner, F.R., Burland, V., Plunkett, G. III, Sofia, H.J., Daniels, D.L. and  
Analysis of the Escherichia coli genome VI: DNA sequence of the  
region from 92.8 through 100 minutes  
Nucleic Acids Res. 23 (12), 2105-2119 (1995)  
JOURNAL MEDLINE 95334362  
PUBMED 7610040  
TITLE 2 (bases 1 to 338534)  
AUTHORS Plunkett, G. I. I.  
DIRECT SUBMISSION  
SUBMITTED (22-AUG-1994) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459

COMMENT  
On Apr 17, 1996 this sequence version replaced gi:536929.  
This sequence was determined as part of the E. coli Genome Project  
(Frederick R. Blattner, director) at the University of  
Wisconsin-Madison. Supported by award HG00301 from the NIH Human  
Genome Project. The entire sequence was independently determined  
from E. coli MG1655; overlaps with other sequence determinations  
are annotated. The start of this entry overlaps the end of the  
entry ECOW89 (U00006) by 1885 bp.

FEATURES  
source

Location/Qualifiers  
1..338534  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K-12"  
/sub\_strain="MG1655"  
/db\_xref="taxon:562"  
/map="92.8 to 100 minutes"  
/note="This sequence comprises the following lambda  
clones: DD928(EC27-278), DD930(EC30-188),  
DD933(EC27-1320), DD935(EC30MM32), DD937(EC27-1070),  
DD941(EC17-142), DD945(EC17-8), DD947(EC248-34),  
DD949(EC22-169), DD952(EC27-297), DD953(EC27-409),  
DD956(EC17-103), DD958(EC30MM1), DD960(EC30MM2),  
DD962(EC21-104), DD965(EC23A-40), DD968(EC30K60A-49p),  
DD970(EC19-202), DD974(EC19-61), DD975(EC18-233),  
DD977(EC30MM7), DD980(EC27-1151), DD981(EC18-3),  
DD984(EC17-101), DD987(EC18-115), DD990(EC17-136),  
DD992(EC18-282), DD995(EC17-115), DD997(EC18-113); M13mp19  
or Janus vectors were used for subcloning"  
1..1885  
/note="1885 bp overlap with end of Genbank Accession  
Number U00006 (ECOW89)"  
complement(1..398)  
/note="ORF\_f326 of Genbank Accession Number U00006  
(ECOW89)"  
/codon\_start=1  
/label="ORF\_f32p  
/protein\_id="AA96985.1"  
/db\_xref="GI:1263173"  
/translation="MGFTTRVKSSEAKKPFNFALFMDKRYGTFILAIIVAFGLSP  
EYFLTNNTQIFVSSVVLIGMGEFFAILVAGIDLVSQALIALSGMTAKMLAGV  
DPEFLAMIGVLVGMALINGCLVWVTL"  
277..397  
/note="predicted bend of 75.10 degrees"  
complement(377..1909)  
/note="ORF\_f510"  
/codon\_start=1  
/label="ORF\_f510  
/protein\_id="AA96986.1"  
/db\_xref="GI:536931"  
/translation="MATPYISMAGIGKSPFPAHALKSVNLTYVPGRIHALGENGAK  
STLMKVLGSIHEPTKGTITINNINYNKLDHKAALGIGIYOELSVIDELTYENLY  
IGRHLTKIIGVNIIDREMRVRAAMLRVGLKVDLDEKVNLSISHOMEIAITL  
MDAYVITMDREPTSLTNKVDYLYPLNOLKREGATYVISHKLAETIRICDRTYVM  
KGGSYSGCIVSDVSHDDIVRLMVGRELONRRPNAKENVSNLAHETVFEVRYNTSRDR  
KVVRLDISFVSCGELIGFAGLVSGRTLELNCILPVDRKAGEIRLNGKDISPRLD  
AVKQMAVITTESRNDGFPNFSIAQNMVAISRLKDGKYGAMGLFHEVDERTAEQ  
RELIALKCHSVNQNTTELSGNGQKVLISKMLCCPEVLIIPDEPTGIDVGAQAEIYK  
VMRLADGKVLVMSSELPEIITVCDRIAVPCEGRITQILNTRDMSEETIWMALP  
QE"

misc\_structure  
CDS

misc\_structure  
misc\_structure  
promoter  
repeat\_region  
CDS  
809..931  
/note="predicted bend of 74.75 degrees"  
1768..1889  
/note="predicted bend of 75.87 degrees"  
complement(1983..2010)  
/note="promoter matrix score of 55; putative; resembles  
flh-regulated promoter"  
complement(1985..2021)  
/standard\_name="RBP; repetitive extragenic palindromic  
element"  
/note="contains 1 REP sequence"  
complement(2036..2971)  
/note="ORF\_f311"  
/codon\_start=1  
/label="ORF\_f311  
/protein\_id="AA96987.1"  
/db\_xref="GI:536932"  
/translation="MNYLKYPSGTLVGLMSTSAFAAEVAVVLTLSNPVMDKK  
GIEDEAKTIGVSVDIFASPSGDFOQOLDFEDLSNKNKYGIAFAPLSSVNLVMPVAR

```
AMKGIYLVNLDKIDMDNLKKAQNVFAVTTDNVAVGAKGAPFIIDLAGEGEVA
IIEGKAGNSGEARNGATEAPFKASQIKLIVASOPADMDRIKALDVATNTVLRNPNIK
AIYCAANDTMMAGVAOAVANAGKTGKLVJCTGCTGIPBARKMVEAGQMTATVAVQNPADIG
ATGKLKLVDAEKSQKVIPLDKABEFLVUSILVTO"
misc_structure
3010..3121
/!note="predicted bend of 73.22 degrees"
CDS
complement(3030..3953)
/!note="ORF f307"
/codon_start=1
/transl_table=11
/!label=ORF f307
/protein_id="AAA96988.1"
/db_xref="GI:536933"
/translation="MASSIQGQENVMSQSEPDALPNGIGHAPYLRMKQEGMTENESR
IYEWLKPENLSCAPAIKDVAEALAVSEAMIVKSKLIGFSGFNRRLBALDYSQSE
OVLPSBELADVEAPQDVNVKFNITLRTIMEGOSIVNVDIEHRAAFQARQRLIGA
GGSNACADVQHKFPIRGVCOAYPDAMMMASILLQEGVIVTVTHSGTSVKA
VELAKNGAKITICITHSYHSPAKIADVIICSRAETPLIGRNASARILQVTLIDAF
VSAQONTBOANINWOKGAIYDPPSPQALK"
complement(3952..3979)
promoter
/!note="promoter matrix score of 61; putative"
misc_structure
4151..4251
/!note="predicted bend of 72.06 degrees"
CDS
4279..4728
/!gene="yjcA"
/codon_start=1
/transl_table=11
/!label=ORF O149
/protein_id="AAA96989.1"
/db_xref="GI:536934"
/translation="MKKIAFCCHVIFILKHEIVAHIVERGVEIDKGTWSSERTDYP
HYASQVALAVAGVDGIIICITGVGISTIAANKFQIRAVVCSSEYASQLSRQNDT
NYLAFGRVGLAEAKMTVDAMGAQYEGGRHQORVEAIVAIQRRN"
complement(4520..16191)
misc_feature
/!note="corresponds to Genbank Accession Number D90227
(ECOPHN) 1..11672; the related sequence Genbank Accession
Number J05260 (ECOPHNAO) is from E. coli B, not K-12"
complement(4739..5146)
CDS
/!gene="phnQ"
/complement(4739..5146)
/!gene="phnQ"
/!note="annotated in Genbank Accession Number J05260
(ECOPHNAO), but probably not a gene"
/codon_start=1
/transl_table=11
/!protein_id="AAA96990.1"
/db_xref="GI:536935"
/translation="MPAPYTSPIILIFHSPLFIQLPTLTIKALAIYOLPAATFKL
AAAVLISAPIVAHFTLPAVARIFVSPFTVAVIVNRLNLIANGVVDNIGLRNVD
SVSRVKNRNSGRKFIQPOKETLSTLIITHPK"
complement(5272..6030)
CDS
/!gene="phnP"
/complement(5272..6030)
/!gene="phnP"
/codon_start=1
/transl_table=11
/!protein_id="AAA96991.1"
/db_xref="GI:536936"
/translation="MSLTLTLTGAGAGVAVNGCECAACARARSPOYRRQSGGVV
KENDAITLADGLHDLADRWSPGSGFOQLTIHMDHVOGLFPLRWGSDPIPYGSP
DEQGGDILFKHGLDPSHTYBPFVVPDGLQVPLPLNHSKLTFGYLETASRYA
WLSPIAGIPKTIKFLRNNOPOVMMDCSHPRADAPRNHCDLNTVLAHQVIRSPRV
ILTHISHOPDMLNENALPSGEVGFDMGEIGA"
5452..5586
misc_structure
/!note="predicted bend of 75.60 degrees"
CDS
complement(6032..6466)
/!gene="phnO"
/complement(6032..6466)
/!gene="phnO"
/codon_start=1
/transl_table=11
```

```
/protein_id="AAA96992.1"
/db_xref="GI:536937"
/translation="MPACELRPATQYDVAVALICELKQAEFDHHAFFVGNANLRD
PNMRYHLALDGSVVMGILHLOFHLHVNVMGEILOELVVMQOARGLVNGSKLMAAE
BEARQGAENTELISTVVKRHDARFYLREBGYSQSHRFYKAL"
complement(6453..7010)
/!gene="phnM"
/complement(6453..7010)
/!gene="phnM"
/!note="HisP-like nucleotide binding protein (ATP-binding
protein)"
/codon_start=1
/transl_table=11
/!protein_id="AAA96993.1"
/db_xref="GI:536938"
/translation="MMGKILIMGSGSGKDSILAEILRROTOLLVAHRYITDASA
GSENHIALSEORFPTRAONULASWHANGLYGVGVIDIMLAGFDVLVNGSRHL
POARARYOSALIPVCLQVSPETLRQRLNRGRNENASFINARLAARYTPDCHTLNN
DGLRSQVDTLTIHQEKHACL"
complement(7010..8146)
/!gene="phnM"
/complement(7010..8146)
CDS
Query Match 100.0%; Score 927; DB 1; Length 338534;
Best Local Similarity 100.0%; Pred. No. 3.5e-264;
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATGTTACGCCGATCATGTTAATTTGACAGATGATCAGCGAGTGCAGGACTTTTGT 60
DB 331239 GTCATGTTACGCCGATCATGTTAATTTGACAGATGATCAGCGAGTGCAGGACTTTTGT 331180
QY 61 ACTTCCTGTTTCGATTAAGTTAGTGGCAATTTAGTACGAAACATGCAACCCGACATTC 120
DB 331179 ACTTCCTGTTTCGATTAAGTTAGTGGCAATTTAGTACGAAACATGCAACCCGACATTC 331120
QY 121 TATCGTTGAAGACGAGTGTGTAACACGCAACACGTTGAAAAGTATTTTGAACCGGAAG 180
DB 331119 TATCGTTGAAGACGAGTGTGTAACACGCAACACGTTGAAAAGTATTTTGAACCGGAAG 331060
QY 181 CTATGATGTTTTGCAACGACAGATGGCGCGGAAATGATCAGATCCTCTGTAATATGA 240
DB 331059 CTATGATGTTTTGCAACGACAGATGGCGCGGAAATGATCAGATCCTCTGTAATATGA 331000
QY 241 CATCAACCTGATGATCATGATATCATCGCGGGTAAAGAGCTTCTGTAGCGCG 300
DB 330999 CATCAACCTGATGATCATGATATCATCGCGGGTAAAGAGCTTCTGTAGCGCG 330940
QY 301 TGAACCTGCGGACGACGCGAATGTTGCTGATGTTCTGACTGGCGGTGAACAAGT 360
DB 330939 TGAACCTGCGGACGACGCGAATGTTGCTGATGTTCTGACTGGCGGTGAACAAGT 330880
QY 361 CGATTAATAATCTCGGCGCTGGAATCGGTGACATGATCATCAACCAACCGTTCAACCC 420
DB 330879 CGATTAATAATCTCGGCGCTGGAATCGGTGACATGATCATCAACCAACCGTTCAACCC 330820
QY 421 GCGTGAATCGATTCGTGACGCAACCTACTGTCCTGACCATGAATCTGGTACTGT 480
DB 330819 GCGTGAATCGATTCGTGACGCAACCTACTGTCCTGACCATGAATCTGGTACTGT 330760
QY 481 CAGCGAAGAACGTCGAGCGTTGAAAGCTACAACTCAATGTTGGAACTGACATCA 540
DB 330759 CAGCGAAGAACGTCGAGCGTTGAAAGCTACAACTCAATGTTGGAACTGACATCA 330700
QY 541 CAGCGTGTGTTGATCGGCGCTGATGAGGACGACGATCAAGCTGCGCGGACGAGTTCCG 600
DB 330699 CAGCGTGTGTTGATCGGCGCTGATGAGGACGACGATCAAGCTGCGCGGACGAGTTCCG 330640
QY 601 CGCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGCTGAATGCTGAA 660
DB 330639 CGCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGCTGAATGCTGAA 330580
QY 661 GAAAAAGACCGGCGGTGATGAACCGCAAGACGATCTGTACAGTGAACATCGCCCG 720
```

Db	330579	GAAATGACGGCGGTGAGTGAACACGCACGACCGTACTGTAGACGTGACGATCCGCCG	330520
Oy	721	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	780
Db	330519	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	330460
Oy	781	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	840
Db	330459	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	330400
Oy	841	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	900
Db	330399	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	330340
Oy	901	GTCGCTGTAACTTCTTACTGCGCTT	927
Db	330339	GTCGCTGTAACTTCTTACTGCGCTT	330313
RESULT 4 LOCUS DEFINITION ACCESSION KEYWORDS ORGANISM REFERENCE AUTHORS	AE005671/c Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 290 AE005671 AE005174 AE005671.1 GI:12519428  Escherichia coli O157:H7 EDL933 Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  1 (bases 1 to 6794) Perna,N.T., Plunkett,G. III, Burdand,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Groebbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Weich,R.A. and Blattner,F.R.		
TITLE JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 Nature 409 (6819), 529-533 (2001) 21074935 11206551  2 (bases 1 to 6794) Perna,N.T., Plunkett,G. III, Burdand,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Groebbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Weich,R.A. and Blattner,F.R.		
TITLE JOURNAL FEATURES SOURCE	TITLE Direct Submission Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  Location/Qualifiers 1..6794 /organism="Escherichia coli O157:H7 EDL933" /mol_type="genomic DNA" /strain="EDL933" /serotype="O157:H7" /db_xref="caxon:155864" /note="enterohemorrhagic" 136..609 /gene="crea" /name="synonym: Z6000" 136..609 /gene="crea" /name="synonym: Z6000" /function="orf, Unknown function" /note="Residues 1 to 157 of 157 are 100.00 pct identical to residues 1 to 157 of 157 from Escherichia coli K-12 Strain MG1655: B4397" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAG59577.1"		
CDS	gene	gene	CDS
Db	330579	GAAATGACGGCGGTGAGTGAACACGCACGACCGTACTGTAGACGTGACGATCCGCCG	330520
Oy	721	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	780
Db	330519	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	330460
Oy	781	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	840
Db	330459	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	330400
Oy	841	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	900
Db	330399	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	330340
Oy	901	GTCGCTGTAACTTCTTACTGCGCTT	927
Db	330339	GTCGCTGTAACTTCTTACTGCGCTT	330313
RESULT 4 LOCUS DEFINITION ACCESSION KEYWORDS ORGANISM REFERENCE AUTHORS	AE005671/c Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 290 AE005671 AE005174 AE005671.1 GI:12519428  Escherichia coli O157:H7 EDL933 Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  1 (bases 1 to 6794) Perna,N.T., Plunkett,G. III, Burdand,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Groebbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Weich,R.A. and Blattner,F.R.		
TITLE JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7 Nature 409 (6819), 529-533 (2001) 21074935 11206551  2 (bases 1 to 6794) Perna,N.T., Plunkett,G. III, Burdand,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Groebbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Weich,R.A. and Blattner,F.R.		
TITLE JOURNAL FEATURES SOURCE	TITLE Direct Submission Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  Location/Qualifiers 1..6794 /organism="Escherichia coli O157:H7 EDL933" /mol_type="genomic DNA" /strain="EDL933" /serotype="O157:H7" /db_xref="caxon:155864" /note="enterohemorrhagic" 136..609 /gene="crea" /name="synonym: Z6000" 136..609 /gene="crea" /name="synonym: Z6000" /function="orf, Unknown function" /note="Residues 1 to 157 of 157 are 100.00 pct identical to residues 1 to 157 of 157 from Escherichia coli K-12 Strain MG1655: B4397" /codon_start=1 /transl_table=11 /product="orf, hypothetical protein" /protein_id="AAG59577.1"		
CDS	gene	gene	CDS
Db	330579	GAAATGACGGCGGTGAGTGAACACGCACGACCGTACTGTAGACGTGACGATCCGCCG	330520
Oy	721	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	780
Db	330519	TATTCTGTAACATTTTCGAATCTAGCCGGCATACGCCGGAATCATCGCACCATTCACGG	330460
Oy	781	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	840
Db	330459	TGAAGGTATATGCGCTCTGCGGATCTGGAAGATTAATCGGCTTACCACCGTCAAAAAA	330400
Oy	841	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	900
Db	330399	AACGCGCGCTTTTAGCGCGCGCTTTTATTTTCAACCTATTTCAGATACGTAATCTATC	330340
Oy	901	GTCGCTGTAACTTCTTACTGCGCTT	927
Db	330339	GTCGCTGTAACTTCTTACTGCGCTT	330313
RESULT 4 LOCUS DEFINITION ACCESSION KEYWORDS ORGANISM REFERENCE AUTHORS	AE005671/c Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 290 AE005671 AE005174 AE005671.1 GI:12519428  Escherichia coli O157:H7 EDL933 Escherichia coli O157:H7 EDL933 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  1 (bases 1 to 6794) Perna,N.T., Plunkett,G. III, Burdand,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Groebbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Weich,R.A. and Blattner,F.R.		
TITLE JOURNAL MEDLINE PUBMED			

CDS  
complement (4205..4921)  
/gene="arcs"  
/function="regulator; Global regulatory functions"  
/note="Residues 1 to 238 of 238 are 100.00 pct identical to residues 1 to 238 of 238 from *Escherichia coli* K-12 Strain MG1655: B4401"  
/codon\_start=1  
/transl\_table=1  
/product="negative response regulator of genes in aerobic pathways" (genes, arcs and cpsA)"  
/protein\_id="AAG59581.1"  
/db\_xref="GI:12519433"  
/translation="MOTPHILVEDELVTRNTLKSIFAEQDYFEATGEMHOIIS  
EYDINLVMDINLPGKNGLLARELBOQVALMFLTGDRNEVDKILGELGADYIT  
KPFNPRLTIRARNLISRTMNLGTVSEERSVSEYKFKGMEIDINSRELIPDEQYK  
LPRSEFRMLHFCENPKIOSRAELKKMTGRELKPHRTYDVITIRIRKHFESTPDT  
PEIATIHGEGRFCGDLED"  
5017..5157  
/gene="yjjy"  
/note="synonym: Z6005"  
5017..5157  
/gene="yjjy"  
/function="orf; Unknown function"  
/note="Residues 1 to 46 of 46 are 100.00 pct identical to MG1655: B4402"  
/codon\_start=1  
/transl\_table=1  
/product="orf; hypothetical protein"  
/protein\_id="AAG59582.1"  
/db\_xref="GI:12519434"  
/translation="MTKVRNCVLDALSIINNNIISLVGTFPODPVSKTAVITLIT  
AT"  
5557..6243  
/gene="last"  
/note="synonym: Z6006"  
5557..6243  
/gene="laet"  
/function="orf; Unknown function"  
/note="Residues 1 to 228 of 228 are 97.80 pct identical to MG1655: B4403"  
/codon\_start=1  
/transl\_table=1  
/product="orf; hypothetical protein"  
/protein\_id="AAG59583.1"  
/db\_xref="GI:12519435"  
/translation="MRTITIIIVAPARAENITGAADAMKTGFSELRIVDSQAHLEPAT  
RWYHGSGLIDINIKVPTLASLHDVDTVAITARSRAKHYVATPELVLEKES  
SMMSHALYVGRBDSGLTNEELADLVDPVADYSINXGAIVXYCYOLAILIO  
OPTKSDTADQHQLOALRERAMALLITLAVADDIKLVDPWLOQRIGLLQRDPMHLRL  
LHDIENKIRK"  
ORIGIN  
Query Match 99.3%; Score 920.6; DB 1; Length 6794;  
Best Local Similarity 99.6%; Pred. No. 1.3e-262;  
Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 1 GTCAATTAACGCCGATCATGTAATTTGACATCATCAGCGAGTGAAGGACTTTTGT 60  
DB 5021 GTCAATTAACGCCGATCATGTAATTTGACATCATCAGCGAGTGAAGGACTTTTGT 4962  
QY 61 ACTTCCTGTTGATTTAGTTAGTGAACAAATGACACCCCGACATTTCT 120  
DB 4961 ACTTCCTGTTGATTTAGTTAGTGAACAAATGACACCCCGACATTTCT 4902  
QY 121 TATCGTTGAAGAGAGTGTGTAACGACACGTTGAAAGTATTTTCCGAAGG 180  
DB 4901 TATCGTTGAAGAGAGTGTGTAACGACACGTTGAAAGTATTTTCCGAAGG 4842  
QY 181 CTATGATTTTCCGAAGGACAGATGCGCGGAAATGCATCCTCTCTGAATATGA 240  
DB 4841 CTATGATTTTCCGAAGGACAGATGCGCGGAAATGCATCCTCTCTGAATATGA 4782

QY 241 CATCAACCTGTGATCATGTGATATCAATCTGCCGGTGAAGAACGATCTTCTGTTAGCGG 300  
DB 4781 CATCAACCTGTGATCATGTGATATCAATCTGCCGGTGAAGAACGATCTTCTGTTAGCGG 4722  
QY 301 TGAATCTGCGGACGAGCGCAATGTTGCTGATGTTCCGATCGGCGCGGACAGCAAGT 360  
DB 4721 TGAATCTGCGGACGAGCGCAATGTTGCTGATGTTCCGATCGGCGCGGACAGCAAGT 4662  
QY 361 CGATTAATTTCTGCGCTGCAATTCGTCGATGATGATCAATCACCACCGTTCAACC 420  
DB 4661 CGATTAATTTCTGCGCTGCAATTCGTCGATGATGATCAATCACCACCGTTCAACC 4602  
QY 421 GCGTGAATGACGATTCGTCGACGCAACTATCGTCCGTCACCAAGTATTTGGTACTGT 480  
DB 4601 GCGTGAATGACGATTCGTCGACGCAACTATCGTCCGTCACCAAGTATTTGGTACTGT 4542  
QY 481 CAGCGAAGACGTCGTAAGCTTGAAGAGTCAAGTTCAATGTTGGGAATGACATGAA 540  
DB 4541 CAGCGAAGACGTCGTAAGCTTGAAGAGTCAAGTTCAATGTTGGGAATGACATGAA 4482  
QY 541 CAGCGTGTGATCGGCGCTGATGGGACGATGACGACGCTCCGCGACGAGTTCCG 600  
DB 4481 CAGCGTGTGATCGGCGCTGATGGGACGATGACGACGCTCCGCGACGAGTTCCG 4422  
QY 601 CGCCATGCTTCACTTCTGTGAAACCCAGCGCAATTCAGTCCGCTGAACTGTGAA 660  
DB 4421 CGCCATGCTTCACTTCTGTGAAACCCAGCGCAATTCAGTCCGCTGAACTGTGAA 4362  
QY 661 GAAATGACCGCGGTGAGTGAACCGGACGACCGTACTGTGACGTGACGATCCGCG 720  
DB 4361 GAAATGACCGCGGTGAGTGAACCGGACGACCGTACTGTGACGTGACGATCCGCG 4302  
QY 721 TATTCGTAAACATTTGCAATCTACCGCGGATACCGCGGAATATATCGCCACATTCACG 780  
DB 4301 TATTCGTAAACATTTGCAATCTACCGCGGATACCGCGGAATATATCGCCACATTCACG 4242  
QY 781 TGAAGTTATGCTTTCGCGGATATCTGGAAGTAAATTCGCTTTACACCGTCAAAAA 840  
DB 4241 TGAAGTTATGCTTTCGCGGATATCTGGAAGTAAATTCGCTTTACACCGTCAAAAA 4182  
QY 841 AACGGCGCTTTTGAACCGCGCTTTTATTTTCAACCTTATTTCCAGATCGTAACATC 900  
DB 4181 AACGGCGCTTTTGAACCGCGCTTTTATTTTCAACCTTATTTCCAGATCGTAACATC 4122  
QY 901 GTCCGTGTACTCTTTTACTGCTTT 927  
DB 4121 GTCCGTGTACTCTTTTACTGCTTT 4095

RESULT 5  
AP002569/c 168764 bp DNA linear BCT 20-MAR-2004  
LOCUS *Escherichia coli* O157:H7 DNA, complete genome, section 20/20.  
DEFINITION AP002569 BAO00007  
ACCESSION AP002569.1 GI:13364704  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ishii, K., Hattori, M., Tatsu, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, K., and Shinagawa, H.  
Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7 derived from the Sakai outbreak  
Genes Genet. Syst. 74 (5), 227-239 (1999)  
JOURNAL MEDLINE PUBMED  
20198780 10734605  
REFERENCE 2

[illegible]

CDS  
4284..4700  
/gene="ECs5232"  
/note="similar to YUGD\_ECOLI g11790704 percent identity  
100 in 138 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAB38655.1"  
/db\_xref="GI:13364710"  
/translation="MANPEOLFQREERRLIEELBEGSPDALYTTIENHLSADDE  
LLEKAVAFKIGVETPEBELFEDGIVICCDLISCALNADLIDAQVQLMTLAE  
KRDVYDMGTFFEDPNESGDDEDFVEDDDGVNH"  
complement(4746..5039)  
/gene="ECs5233"  
complement(4746..5039)  
/note="similar to B4256\_ECOLI g11790705 percent identity  
93 in 97 aa (Conserved in E.coli K-12)"  
/codon\_start=1  
/evidence=not\_experimental  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="BAB38656.1"  
/db\_xref="GI:13364711"  
/translation="MVGSGGIAPLAGESDICEIQMYFLPAIRGKLAKKALIKAME  
EARMGKRCLEETFAFLKEATIGLYEHLGFQHIDYALGCTGHVDCVRLREL"  
5442..6638  
/gene="ECs5234"  
5442..6638  
/gene="ECs5234"  
g11267367 percent identity 96 in 389 aa (Conserved in

Query Match  
Best Local Similarity 99.3%; Score 920.6; DB 1; Length 168764;  
Matches 923; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GTGATGTTAGCGCGATCATGTTAATTTGCGAGCATGATCAGGAGTCAAGGACTTTTGT 60  
Db 167518 GTGATGTTAGCGCGATCATGTTAATTTGCGAGCATGATCAGGAGTCAAGGACTTTTGT 167459  
QY 61 ACTTCCGTTTGCATTTAGTTGGCAATTAGTAGCAAAATGATGACACCCCGCATTTCT 120  
Db 167458 ACTTCCGTTTGCATTTAGTTGGCAATTAGTAGCAAAATGATGACACCCCGCATTTCT 167399  
QY 121 TATCGTTGAAGAGAGTGGTGTACACGCAACACGTTGAAAGTATTTTGAAGCGGAGG 180  
Db 167398 TATCGTTGAAGAGAGTGGTGTACACGCAACACGTTGAAAGTATTTTGAAGCGGAGG 167339  
QY 181 CTATGATGTTTTCGAAGCGACAGATGCGCGGAAATGATCATGATCTCTCGAATTTGA 240  
Db 167338 CTATGATGTTTTCGAAGCGACAGATGCGCGGAAATGATCATGATCTCTCGAATTTGA 167279  
QY 241 CATCAACCTGGTATGATGATATCAATTCGCGGGTAAAGACGCTTCTTTAGCGCG 300  
Db 167278 CATCAACCTGGTATGATGATATCAATTCGCGGGTAAAGACGCTTCTTTAGCGCG 167219  
QY 301 TGAACCTGCGGACGACGATGTTGCGTGTATGTTCCGACTGCGCGGTGACCAAGAGT 360  
Db 167218 TGAACCTGCGGACGACGATGTTGCGTGTATGTTCCGACTGCGCGGTGACCAAGAGT 167159  
QY 361 CGATTAATTTCTGGCCTCGAAATGCGTGCAGATGACTATCAACCAACCGTTCAACCC 420  
Db 167158 CGATTAATTTCTGGCCTCGAAATGCGTGCAGATGACTATCAACCAACCGTTCAACCC 167099  
QY 421 GCGTGAACCTGACGATTCGTGACGCAACCTACTGCTCCCGTACCAATTCGGTACTGT 480  
Db 167098 GCGTGAACCTGACGATTCGTGACGCAACCTACTGCTCCCGTACCAATTCGGTACTGT 167039  
QY 481 CAGCGAAGAACGTCGTACGTTGAAAGCTACAAAGTTCAATGTTGGGAACTGACATCAA 540  
Db 167038 CAGCGAAGAACGTCGTACGTTGAAAGCTACAAAGTTCAATGTTGGGAACTGACATCAA 166979

QY 541 CAGCGCTTCTGTATGATCGCGCCCTGATGAGCGACATCAAGCTGCCCGGACGAGTTCCG 600  
Db 166978 CAGCGCTTCTGTATGATGATCGCGCCCTGATGAGCGACATCAAGCTGCCCGGACGAGTTCCG 166919  
QY 601 CGCCATGCTTCACTTTGTGTAAACCAGGCAAAATTCAGTCCCGTGAATGCTGAA 660  
Db 166918 CGCCATGCTTCACTTTGTGTAAACCAGGCAAAATTCAGTCCCGTGAATGCTGAA 166859  
QY 661 GAAATGACCGGCGGAGCTGAAACCGGACACCGTATCTGAGACGTAGATCCGCCG 720  
Db 166858 GAAATGACCGGCGGAGCTGAAACCGGACACCGTATCTGAGACGTAGATCCGCCG 166799  
QY 721 TATTCGTAAACATTTTGAATCTACCGCGGATACCGCGAAATCATCGCACCATTCACGG 780  
Db 166798 TATTCGTAAACATTTTGAATCTACCGCGGATACCGCGAAATCATCGCACCATTCACGG 166739  
QY 781 TGAAGTTATCGCTTCTGCGGATCTGGAAGATTATCGGCTTTACCAACGTCMAAAA 840  
Db 166738 TGAAGTTATCGCTTCTGCGGATCTGGAAGATTATCGGCTTTACCAACGTCMAAAA 166679  
QY 841 AACGGGCTTTTTCAGCGCGCTTTTATTTTCAACCTTATTTCCAGATGATCATC 900  
Db 166678 AACGGGCTTTTTCAGCGCGCTTTTATTTTCAACCTTATTTCCAGATGATCATC 166619  
QY 901 GTCCGTTGATCTTCTTACTGCTTT 927  
Db 166618 GTCCGTTGATCTTCTTACTGCTTT 166592

RESULT 6  
AE016772/c 86698 bp DNA linear BCT 24-DEC-2002  
LOCUS  
DEFINITION  
Escherichia coli CFT073 section 18 of 18 of the complete genome.  
ACCESSION  
AE016772 AE014075  
VERSION  
AE016772.1 GI:26111641  
KEYWORDS  
ORGANISM  
SOURCE  
Escherichia coli CFT073  
Escherichia coli CFT073  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 (bases 1 to 86898)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Heckert,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.  
Extensive Mosaic Structure Revealed by the Complete Genome Sequence  
of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
12471157  
2 (bases 1 to 86898)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Heckert,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.  
Direct Submission  
Submitted (20-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
1..86898  
LOCATION/Qualifiers  
1..86898  
/organism="Escherichia coli CFT073"  
/mol\_type="genomic DNA"  
/strain="CFT073"  
/db\_xref="taxon:199310"  
complement(152..1495)  
/gene="gntP"  
/locus\_tag="c5401"  
complement(152..1495)  
/gene="gntP"  
/locus\_tag="c5401"  
/function="transport; Transport of small molecules:  
carbohydrates, organic acids, alcohols"  
/note="Gluconate permease 3; Escherichia coli K-12



ortholog: b4321; Escherichia coli O157:H7 ortholog: z5919"  
/codon\_start=1  
/transl\_table=11  
/product="High-affinity gluconate transporter"  
/protein\_id="AA083823.1"  
/db\_xref="GI:26111642"  
/translation="MHVNLIVYFGIGLMLVNLKFKINSVALLVLAALVSGMLAGM  
DMLSLHTKAGFGNTLGLBLAIVFGAVGKLVDSGAHQIAPHTLARLGLRVQI  
SVIIGLIFGLAMFYEVAFIMLAPLIVIAAEAKIPPLKLAIPVAAPATTAHLPPO  
PSPVALVNAVAGDMGVYIYGVLTIVISVACGLIPKFLGNLERPTSPPLAKDOVD  
MNNLPSEGVSLIVPLIPATIMSTTIANIWLVDTPAVENVFNGSSPLAMFIAWVA  
FVLFGTARGHDQMWNKAPESAKSIAMIILIIIGAGVUKOTIIDTGIGDTIGLMSH  
GNISPIYIMALITVLIRLATGQVSAMTLAGIISAIIIDPATGOLGVAVPALVAT  
AAGSNTLTHINDSFWLFRGYFDLSVDTLKTGLLELVNSVGLIIVILISMVA"  
1835..3019  
/gene="uxuA"  
/locus\_tag="c5402"  
1835..3019  
/locus\_tag="uxuA"  
/gene="uxuA"  
/locus\_tag="c5402"  
/EC\_number="4.2.1.8"  
/function="enzyme; Degradation of small molecules: Carbon  
compounds"  
/note="Escherichia coli K-12 ortholog: b4322; Escherichia  
coli O157:H7 ortholog: z5920"  
/codon\_start=1  
/transl\_table=11  
/product="Mannonate dehydratase"  
/protein\_id="AA083824.1"  
/db\_xref="GI:26111643"  
/translation="MEQTRWVGPNDPVSILADVQAQAGVVTALHHIPGEVSE  
ILKRAIVEDAGLVMSQSVIPGHEDIKTHGYEQMIANYQOTLRNLACGRTVCY  
NMPVLVDTRTDELYLPDGSKALRFQDIEFAFEMHILKRPGEADYEESEIAQAB  
RAVMSDEDKARLTRNIAGLPGAEGYTLDPGRKHLYKOIDKAKLENFVFLKA  
IIPVAEVCVRMAVHPDDPRPRICPRIVSTIEDQMWDVTPNSANGFTMTGSGY  
VRAINDLVMIKQFGPRITYFHLRSTRENPETFEAHNLNGVDVMEVKAIVEE  
HRRKAEKEDLIIPRPDGHQMLDLKKKTNPQSAIGRLKGLAEVRGVELAIQRAFP  
SR"  
3100..4560  
/gene="uxuB"  
/locus\_tag="c5403"  
3100..4560  
/gene="uxuB"  
/locus\_tag="c5403"  
/EC\_number="1.1.1.57"  
/function="enzyme; Degradation of small molecules: Carbon  
compounds"  
/note="Escherichia coli K-12 ortholog: b4323; Escherichia  
coli O157:H7 ortholog: z5921"  
/codon\_start=1  
/transl\_table=11  
/product="D-mannonate oxidoreductase"  
/protein\_id="AA083825.1"  
/db\_xref="GI:26111644"  
/translation="MTTIVDSNLPVAPSPMDHSRLSRIVHLGCGAFHRAQALYTH  
LLESTSDWGCIEVNLPMDNDRVLENLKQOQLYVAEKAESTELKIGSKELH  
PEIDGEGEGLINMAAPOTAIIVLTVEKGYCADPAAGOLDIANNPLIKHLENTLTA  
ATGYIEARLRLREKGLKAPTVMSCDNVRNGVAVAYVLGLAQARDPOLAMIENTV  
TPCTWVDRIVPAPETLOETIADQGVDPQCIACEPRQVTEDFNFGNRDWMKV  
GAQFVADVVPFEMMKLRMLNGSHSLAYIGLYGTYETIADVTNPPYRKAAPALMQE  
QAPLTSPBGTDLNAATLILIERFSPNSLRRHTWQIAMDGSOQLPQRLDPPVYLHQN  
GGSWRILALGVAGMRYTOGVDEQGNALIVDVPMALEPKINAYQGAADRIVKALGLS  
GFADLPQNAQFVAVTAAYQGLCERGARCEVAL"  
4775..5548  
/gene="uxuR"  
/locus\_tag="c5404"  
4775..5548  
/gene="uxuR"  
/locus\_tag="c5404"  
/locus\_tag="uxuR"  
/function="regulator; Degradation of small molecules:  
Carbon compounds"  
/note="Escherichia coli K-12 ortholog: b4324; Escherichia  
coli O157:H7 ortholog: z5922"

/codon\_start=1  
/transl\_table=11  
/product="Uxu operon trancriptional regulator"  
/protein\_id="AA083826.1"  
/db\_xref="GI:26111645"  
/translation="MKSATSAQRPYEVGMIRDLIKTPYNGELPPEPEREIAEMID  
VTRTVREALIMLEIKGLVEVRGAGIYVLDSSGSNNASPDANVCNDGAPFELQAR  
QLIESNIASFALQATREDIYVNRQALQEEERLNASAGSSSGDMQFPLAIAEYTH  
NSMLVELFQSMQHMNSPMWIOHSHLDSDLYRKEMLGDKHKOILAALIKDARAACL  
ANMOHLENKORLREBSNVDTIFYDGYLFDSPNLDKVA"  
5953..6345  
/gene="yjiD"  
/locus\_tag="c5405"  
5953..6345  
/gene="yjiD"  
/locus\_tag="c5405"  
/note="Escherichia coli K-12 ortholog: b4326; Escherichia  
coli O157:H7 ortholog: z5925"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yjiD"  
/protein\_id="AA083827.1"  
/db\_xref="GI:26111646"  
/translation="NRKMRQSVQTVLPSTGNNTLSRDSVCRDLFOUFSRSP  
ILLVSGPMPQGHNSDKLQSVYCRQLSALFLHBPRIALQVNLKAYCHELAISL  
EMMLYHDDPPLTFDLVWQKSGSWHRTWPQ"  
complement(6393..6734)  
/locus\_tag="c5406"  
complement(6393..6734)  
/locus\_tag="c5406"  
/note="Residues 1 to 113 of 113 are 95.57 pct identical to  
residues 424 to 536 of 536 from GenPept.129 :  
>B|AA061895.1| (A0289032) putative  
phosphoenolpyruvate-protein phosphoceryltransferase ppe  
[Escherichia coli]"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein"  
/protein\_id="AA083828.1"  
/db\_xref="GI:26111647"  
/translation="MIEVPAVWIDADKLASEVDFSGTNDLTQYIMAADRGSTYAK  
LVDCNDVAIVINAIWVQAGRNNEIPVSMCGEAGDIIQOTARLLTWGIDKLSPSRL  
PALKAIRTSH"  
complement(6771..7760)  
/gene="yjiB"  
/locus\_tag="c5407"  
complement(6771..7760)  
/gene="yjiB"  
/locus\_tag="c5407"  
/function="putative regulator"  
/note="Escherichia coli K-12 ortholog: b4327; Escherichia  
coli O157:H7 ortholog: z5926"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical trancriptional regulator yjiE"  
/protein\_id="AA083829.1"  
/db\_xref="GI:26111648"  
/translation="MKRKRSLAABWDVCPQGRLTENVYMDGCAVLANIETKMLYD  
FITLEKCRNFQSAVSRVNSQPAFSRRITALLQAGVBLPNROVTPYLSQSOCKITHS  
QIRHLQQLSESLAELRGSDYAQRKIKIAAHSLSGLPSTISQMPPLFTWAIATAI  
DVEEDLRLKREGSCDIFSFHDEDLLEAFPDHI RLPESQLFPVCADEBGEALFALAQ  
PHEPLNYSRNSYMGRLIRTLTRSELSFSYFVSSMSGLKQVALLDCCGIALPREY  
AIOGRISGQLVILNRDELVIRIQAVAYMNTMRNPVAFRPRELEIEBPQ"  
complement(7747..8919)  
/gene="iada"  
/locus\_tag="c5408"  
complement(7747..8919)  
/gene="iada"  
/locus\_tag="c5408"  
/EC\_number="3.4.19.-"  
/function="enzyme; Degradation of proteins, peptides,  
glyco"  
/note="Escherichia coli K-12 ortholog: b4328; Escherichia

coli 0157:H7 ortholog: z5927"  
/codon\_start=1  
/translation="1  
/product="Isoaspartyl dipeptidase"  
/protein\_id="AA083830.1"  
/db\_xref="GI:26111649"  
/translation="MIDYSAAGPTLLIQNALYAPEDRGICDVLINAKKIIIVASNIIPS  
DIYDCTIVDLGSQLICPGFIDQVHLICGGGAGAPPTTPPEVALSRLLBAGITSVG  
LVGTDISIIRHPSBLAKTLRALNEBISAMLTGAVHSRRTTGSVEEDVAITDRTVIG  
VKCAISDRSSAAPDVYHLANMAAESVGLGKXPGVTFPHNGDSKRLQYVYDLEN  
CVPISKLPTVNRNVPLFEQALFARAGTIDITSSIDEVAPABEIGARAVQAGIP  
LARVLISDNGSOPPEDEGNLTHTIIGVAPETTLTQVAVKDYSEFSLDALRPLTS  
SVAGFLNKGKEIIPGNDADILVMTPERLISQVYARKKMKACVKGTFFETA"

Query Match 98.4%; Score 912.6; DB 1; Length 86698;  
Best Local Similarity 99.0%; Pred. No. 5,3e-260;  
Matches 918; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

1 GTGATGTTACGCCGATCATGTTAATTTGAGATGATCATGAGGAGTCAAGGACTTTTGT 60  
85652 GTGATGTTACGCCGATCATGTTAATTTGAGATGATCATGAGGAGTCAAGGACTTTTGT 85593  
61 ACTTCTGTTTGAATTAAGTTAGGCAATTTAGGTAACAACATGACACCCGACATTTCT 120  
85592 ACTTCTGTTTGAATTAAGTTAGGCAATTTAGGTAACAACATGACACCCGACATTTCT 85533  
121 TATCGTTGAAGACGATTGTTAACAACGACACCTTTGAAAAGTATTTTCGAAGCGGAAG 180  
85532 TATCGTTGAAGACGATTGTTAACAACGACACCTTTGAAAAGTATTTTCGAAGCGGAAG 85473  
181 CTATGATTTTTTGAAGACGATGAGCGCGGAATGATCATGATCTCTCTGAATTTGA 240  
85472 CTATGATTTTTTGAAGACGATGAGCGCGGAATGATCATGATCTCTCTGAATTTGA 85413  
241 CATCAACCTGGTGAATGATGATCATGATCTCTCTGGGTAAGAACGCTTTCTGTTAGCGCG 300  
85412 CATCAACCTGGTGAATGATGATCATGATCTCTCTGGGTAAGAACGCTTTCTGTTAGCGCG 85353  
301 TGAATCGCGGACGAGCGGATGTTGCTGATGTTCTGACGCGCGGTGACACGAAAGT 360  
85352 TGAATCGCGGACGAGCGGATGTTGCTGATGTTCTGACGCGCGGTGACACGAAAGT 85293  
361 CGAATAAATTTCTGGGCTCGAATTCGCTGCAATGATGATCATGACCAACCGTTCAACC 420  
85292 CGAATAAATTTCTGGGCTCGAATTCGCTGCAATGATGATCATGACCAACCGTTCAACC 85233  
421 GCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
85232 GCGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85173  
481 CAGCGAAGAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
85172 CAGCGAAGAACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85113  
541 CAGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
85112 CAGCGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85053  
601 CCGCATGCTTCACTTCTGTGAACCCAGGCAAAATTCAGTCCCGTGTGAATGCTGAA 660  
85052 CCGCATGCTTCACTTCTGTGAACCCAGGCAAAATTCAGTCCCGTGTGAATGCTGAA 84993  
661 GAAATGACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720  
84992 GAAATGACCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 84933  
721 TATTTGTTAAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780  
84932 TATTTGTTAAACATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 84873  
781 TGAAGTTATGCTTCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
84872 TGAAGTTATGCTTCTGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 84813

841 AACGCGCTTTTAAAGCGCGTTTAAATTTTCACTTATTTTCAGATACGATACGATCATC 900  
84812 AACGCGCTTTTAAAGCGCGTTTAAATTTTCACTTATTTTCAGATACGATACGATCATC 84753

901 GTCCGTTGTAACCTTTTCTTACTGCTTT 927  
84752 GTCCGTTGTAACCTTTTCTTACTGCTTT 84726

RESULT 7  
ECODYE  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED

COMMENT

Original source text: E.coli K12 DNA, clone pRB52.  
Draft entry and sequence in computer readable form kindly provided  
by R.S.Buxton, 01-AUG-1985 [1].  
Mutation of the dye gene results in sensitivity to dyes, envelope  
protein changes, loss of expression of alkaline phosphatase, and  
reduced transcription of sex factor F genes. The positions of the  
dye and ompR genes on the K12 map are indicative of gene  
duplication and the dye protein is 28% homologous to the OmpR  
protein.  
A potential transcription termination site is located at positions  
906-940 and a ribosome binding site at 161-166. An UPR was found  
on the complementary strand at positions >1542 to 948 (AA at 1541).  
Location/Qualifiers  
1..1468  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
98..814  
/gene="dye"  
98..814  
/gene="dye"  
/codon\_start=1  
/translation="1  
/protein\_id="AAA23718.1"  
/db\_xref="GI:145818"  
/translation="MOTPHILLVEBELYTRNTLKSIFAEAGYDFEATDGAEMHQILS  
EYDINLVITMDINLPGNGGLLAREIREQANALMFLNDRNEVKKILGLIEIGADYIT  
KPPNRELITRANRLISRTNMIGVSEBRREVSYSFYGMEIDINSRSLIPDGEQYK  
LPRSEFRMLHPCENPGKIOSRABELIKKGTREIKPHDRYDVITIRIRKREESTPDT  
PRLIATIHGEGRFCGDLED"

ORIGIN  
699 bp upstream of AccI site.

Query Match 98.4%; Score 912; DB 1; Length 1468;  
Best Local Similarity 99.9%; Pred. No. 3.7e-260;  
Matches 923; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

5 TGTACGCCGATCATGTTAATTTGAGATGATCATGAGGAGTCAAGGACTTTTGTACT 63  
84932 TGTACGCCGATCATGTTAATTTGAGATGATCATGAGGAGTCAAGGACTTTTGTACT 60  
64 TCCTGTTGATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 123  
84932 TCCTGTTGATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 120  
124 CGTGAAGACGAGTTGTTAACAACGATGTTGAAAAGTATTTTGAAGCGGAAGGCTA 183

```

Db      ||||| 121 CGTTGAAGACGATGGTAAACGCAACAGCTTGAAGAAAGATTTTCCAAACCGGAAGGCTA 180
Qy      ||||| 184 TGATGTTTTGGAAGCGACAGATGGCGGGAATGATCAGATCCTCTCTGAATATGACAT 243
Db      ||||| 181 TGATGTTTTGGAAGCGACAGATGGCGGGAATGATCAGATCCTCTCTGAATATGACAT 240
Qy      ||||| 244 CAACCTGGTGATCATGATATCAATCTGCGGGTGAAGAACGGTCTTCTGTTAGCGCGTGA 303
Db      ||||| 241 CAACCTGGTGATCATGATATCAATCTGCGGGTGAAGAACGGTCTTCTGTTAGCGCGTGA 300
Qy      ||||| 304 ACTGCGCGAGCAGCGCAATGTTGCGTTGATGTTCTGACCTGCGCGTGAACAAGTGA 363
Db      ||||| 301 ACTGCGCGAGCAGCGCAATGTTGCGTTGATGTTCTGACCTGCGCGTGAACAAGTGA 360
Qy      ||||| 364 TAAATTTCTGCGCTCGAAATCGGTGAGATGATCAATGACCAAAACCGTTCAACCGCG 423
Db      ||||| 361 TAAATTTCTGCGCTCGAAATCGGTGAGATGATCAATGACCAAAACCGTTCAACCGCG 420
Qy      ||||| 424 TGAATGACGATTCGTGACGCAACCTACTGTCCTGACCATGATCTGGGTACTGTGAG 483
Db      ||||| 421 TGAATGACGATTCGTGACGCAACCTACTGTCCTGACCATGATCTGGGTACTGTGAG 480
Qy      ||||| 484 CGAAGAACGTCGTAAGGCTGAAGCTAACAAGTTCATGTTGGGAAGTGAACATCAACAG 543
Db      ||||| 481 CGAAGAACGTCGTAAGGCTGAAGCTAACAAGTTCATGTTGGGAAGTGAACATCAACAG 540
Qy      ||||| 544 CGCTTGTTGATGAGGCGCTGATGAGCGAGTGAAGCTGCGCGGAGCGAGTTCGCGCG 603
Db      ||||| 541 CGCTTGTTGATGAGGCGCTGATGAGCGAGTGAAGCTGCGCGGAGCGAGTTCGCGCG 600
Qy      ||||| 604 CATGCTTCACTTCTGTGAAACCAGGCAAAATTCAGTCCGCTGTAAGTGTGTAAGAA 663
Db      ||||| 601 CATGCTTCACTTCTGTGAAACCAGGCAAAATTCAGTCCGCTGTAAGTGTGTAAGAA 660
Qy      ||||| 664 AATGACCGGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 723
Db      ||||| 661 AATGACCGGCGCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 720
Qy      ||||| 724 TCGTAAACATTTGGAATCTGACGCGGATGACGCGGAAATATGATGCCACATTCACGGTGA 783
Db      ||||| 721 TCGTAAACATTTGGAATCTGACGCGGATGACGCGGAAATATGATGCCACATTCACGGTGA 780
Qy      ||||| 784 AGGTATCGCTTCTGCGGATGATGGAAGATTAATCGGCTTTTCAACCGTCAAAAAAAC 843
Db      ||||| 781 AGGTATCGCTTCTGCGGATGATGGAAGATTAATCGGCTTTTCAACCGTCAAAAAAAC 840
Qy      ||||| 844 GGGCGTTTTCAGCGCGGTTTATTTTCAACCTTATTTCCAGATACGATCATGCTG 903
Db      ||||| 841 GGGCGTTTTCAGCGCGGTTTATTTTCAACCTTATTTCCAGATACGATCATGCTG 900
Qy      ||||| 904 CGTTGTAACCTCTTACTGCGCTT 927
Db      ||||| 901 CGTTGTAACCTCTTACTGCGCTT 924

```

```

RESULT 8
AE015448/c 11861 bp DNA linear BCT 18-OCT-2002
LOCUS Shigella flexneri 2a str. 301 section 411 of 412 of the complete
DEFINITION
genome.
ACCESSION AE015448 AE005674
VERSION AE015448.1 GI:24054959
KEYWORDS
SOURCE
ORGANISM
Shigella flexneri 2a str. 301
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
1 (bases 1 to 11861)
Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,
Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Yang, G. W.,
Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P.,
Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,

```

```

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
gene
CDS
gene
CDS

```

Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J. Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157  
Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
12384590

2 (bases 1 to 11861)  
Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B., Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y., Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.  
Direct Submission  
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingshi Jie, Xuanwu Qu, Beijing 100052, P.R. China

Location/Qualifiers  
1..11861  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
complement(185..1852)  
/gene="yjjK"  
/locus\_tag="SF4423"  
/gene="yjjK"  
/locus\_tag="SF4423"  
/locus\_tag="SF4423"  
/note="Residues 1 to 555 of 555 are 99 pct identical to residues 1 to 555 of a 555 aa protein from *Bacterichia coli* O157:H7 EDL933 ref: NP\_291004.1"  
/codon\_start=1  
/cranal\_table=11  
/product="putative ATP-binding component of ABC transport system"  
/protein\_id="AA045837.1"  
/db\_xref="GI:24054960"  
/translation="MAQFYVTHRVGVKVPKPRHILKNISLFFPGAKIGVLGANGKSLRLIMAGIDKIEGARPOPIKIGYLPQEPNPHVRESIEAVSEVNAIKRLDEYVALYADPADDEPKLAAGRLSEIIOADGHNIAVORADALRPLMDWKI ANLSGERRRVALICLLLEKPMILDEPTNLDSEAVMLRPLHDFGTVALATHD RYPLDNVAGHILLEDRGCTPWEKYSMLLEKODRLOABOEARRRSIEKLEMY RQGTGROSKGRARLARLELSTEVOKNETNELPIPGPRIGDVAIVSNLRKSYG DBLLIDLSFSLPKGAVIGIIGPNAGKSTLFRMSGQOPSGTTLTETVLAASVD OPRDNDNKTWSEVSGGLDMKIGTILFRMSPRAYVGRNFGVQDGKVGELSGGER GLHLAKLIVGSGNMILDEPTNLDIETRLALEPLAEFGCAWVISHDRWFLDRIA THILYVDEGKVPFEGNTEYEVYKRTLGADALEPKIKYKRLAK"  
2033..4000  
/gene="slt"  
/locus\_tag="SF4424"  
2033..4000  
/gene="slt"  
/locus\_tag="SF4424"  
/note="Residues 2 to 655 of 655 are 99 pct identical to residues 1 to 654 of a 654 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP\_291005.1"  
/codon\_start=1  
/cranal\_table=11  
/product="soluble lytic murein transglycosylase"  
/protein\_id="AA045838.1"  
/db\_xref="GI:24054961"  
/translation="MMYLIHEDALVEKAKQVTRMLLAAGVCLITVSSVARADLDEOR SRYAOIKOAMRNDVOMVOMRGLKDYLYLYLEKROIITDLMOPATVTRNVRAN PTLPARPTOSRFEVNEIARERDMRGLIASEPKPTGTEOCNYYAKMTGQSEAWO GAKELMTKSPNACDKLFSVWRASGKQDPLATYERITLAKAAGTGLVYTLAGMP ADVOTLASAIIISLANNPNTVLFTFARTGATDTFRQMAAVAFASVARQDENRMLIPS LAQAOQLNEDQIOELRDIYAMRLMGNDVDEQAKRDDAIMRSQSTSLIERVRNALG TDDRQGLNTWRLARLPEAKEDKEMRYQADLLERGERAEAKEIILHOLIQQGFYPMV AAORIGEEYELKIDRAPQVNDSTLTQGPMAVRLEIMVNLNTARSEWNLVKSQSK TEOADLARYAFNNQWMDLSVQATIGKMDHLEERPLAIVNLFPRYTSKREIPOSYA MALIAOESAMNRKYSFVGSAGSLMIMETATHTYKMSIPETSSPGQLDDEPTINTI GTSYIQYVYQQRGNRIRFSSAAYNAGPQVTRWLNLSAGRIDAIVAVESIPSETRGY VKNVLAIDAYRYRFGWDKPTLMSATWEMGRY"

gene 4090..4416  
/gene="trpR"  
/locus\_tag="SF4425"  
4090..4416  
/gene="trpR"  
/locus\_tag="SF4425"  
/note="Residues 1 to 108 of 108 are 100 pct identical to residues 1 to 108 of a 108 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_291006.1"  
/codon\_start=1  
/transl\_table=11  
/product="regulator for trp operon and arch; trp repressor"  
/protein\_id="AAN45839.1"  
/db\_xref="GI:24054962"  
/translation="MAQQSPYSAAMAECRHOEMLRPVDLKAYONDIHLPLNMLT PDEREALGTRVRIVEBELRGENMSQRELKNELDAGATITRGSNSLKAAPVLEKOWLEER VLKSD"  
complement(4513..5030)  
/gene="yjiX"  
/locus\_tag="SF4426"  
/pseudo  
complement(4513..5030)  
/gene="yjiX"  
/locus\_tag="SF4426"  
/note="similar to Escherichia coli K12 of, conserved hypothetical protein gi:1790855 (174 aa). BLAST with identity of 95% in 172 aa. This CDS contains a frameshift. The sequence has been checked and is believed to be correct."  
/pseudo  
/codon\_start=1  
/transl\_table=11  
5073..5720  
/gene="gpmB"  
/locus\_tag="SF4427"  
5073..5720  
/gene="gpmB"  
/locus\_tag="SF4427"  
/note="Residues 1 to 215 of 215 are 100 pct identical to residues 1 to 215 of a 215 aa protein from Escherichia coli O157:H7 ref: NP\_313380.1"  
/codon\_start=1  
/transl\_table=11  
/product="phosphoglyceromutase 2"  
/protein\_id="AAN45840.1"  
/db\_xref="GI:24054963"  
/translation="MLQVYLVRHGETONNAERIRIGOSDPLTAKGEQAOAVATRAK ELGITIHISDLGRTRTAEIILAAGCDIIDSRLREINMGVLEKRHIISLITBEEN IRLQLVNGTVGRIIPGESMOELSDRVNALESCHDLPGSRLPLVSHGICLVST ILGLPMAERRLRLRNCISIRVDYQESLWLASGWVETAGDISHIDAPALDELQR"  
complement(5717..6586)  
/gene="rob"  
/locus\_tag="SF4428"  
complement(5717..6586)  
/gene="rob"  
/locus\_tag="SF4428"  
/note="Residues 1 to 289 of 289 are 100 pct identical to residues 1 to 289 of a 289 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_291009.1"  
/codon\_start=1  
/transl\_table=11  
/product="right origin-binding protein"  
/protein\_id="AAN45841.1"  
/db\_xref="GI:24054964"  
/translation="MDQAGITRDLLIMLEGHLDQPLSLDNVAAKAGYSKWLQRMFKD VTGHAGAYIARARLSKSAVALRLTARPIIDIALQVFPDSQQFTTRAFKQAFOTPAL YRSPREMAFIRPRLRGEFTMBEHLIEDPLIGVQSGLSLOISDRHMR LQGBGVMEFTYEGLTGVOERILTVYGTCPMNLIRKKQDIERITPADAKAGRP INRCELPIPRR"  
6797..7270  
/gene="creA"

CDS  
/locus\_tag="SF4429"  
6797..7270  
/gene="creA"  
/locus\_tag="SF4429"  
/note="Residues 1 to 157 of 157 are 100 pct identical to residues 1 to 157 of a 157 aa protein from Escherichia coli O157:H7 EDL933 ref: NP\_291010.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AAN45842.1"  
/db\_xref="GI:24054965"  
/translation="MKYKHLISLSLIMGLPAHERIGSDYTFKAIQDPKIVFA PDDPVKNVTCVSRKAGIKGIGLAEDPSDAISCCQVGP1ELSDRIKNGAQB VYFKRSTSLVPSKSLQVAFPYAKRRAALAYLASDKVVGSPKNAISAVPWPWRQ"  
7283..7972  
/gene="creB"  
/locus\_tag="SF4430"  
7283..7972  
/gene="creB"  
/locus\_tag="SF4430"  
/note="Residues 1 to 229 of 229 are 98 pct identical to residues 1 to 229 of a 229 aa protein from Escherichia coli K12 ref: NP\_418615.1"  
/codon\_start=1  
/transl\_table=11  
/product="catabolic regulation response regulator"  
/protein\_id="AAN45843.1"  
/db\_xref="GI:24054966"  
/translation="MORETVMLVEDDQGIADTLVYMLQDEGFAVEFERGLPYLDAR QQVDPWILDVGLPDISGRFLCQOLALHPLVPLTARSEVAVRLLEIGADYV AKPSPREVCARVATLRLRVKFTSPVIRIGHFELNPAQIISFDPPLTRREF LLAKTLKSPGRVWMSRQQLMDSVWEDADDTYRTVDIHIKTRAKLRALNPLSPINT HRGMYSLRGL"  
7972..9396  
/gene="creC"  
/locus\_tag="SF4431"  
7972..9396  
/gene="creC"

Query Match 98.2%; Score 910.2; DB 1; Length 11861;  
Best Local Similarity 99.6%; Pred. No. 1.9e-259;  
Matches 923; Conservative 0; Mismatches 3; Indels 1; Gaps 1;  
QY 1 GTCATGTTACGCCGATCATGTTAATTTGACATGATCATCAGCAGCTTGTGT 60  
Db 11681 GTCATGTTACGCCGATCATGTTAATTTGACATGATCATCAGCAGCTTGTGT 11622  
QY 61 ACTTCTGTTTGCATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 120  
Db 11621 ACTTCTGTTTGCATTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT 11562  
QY 121 TATCGTTGAAGAGAGTTGTTAAGCGAAGCAACGTTGAAAGTATTTTGAAGCGAAG 180  
Db 11561 TATCGTTGAAGAGAGTTGTTAAGCGAAGCAACGTTGAAAGTATTTTGAAGCGAAG 11502  
QY 181 CATATGTTTTCGAAGCAGATGCGCGGAATCATCATGATCTCTGTAATTTGA 240  
Db 11501 CATATGTTTTCGAAGCAGATGCGCGGAATCATCATGATCTCTGTAATTTGA 11442  
QY 241 CATCAACCTGGTATCATGATCATCTCCGGGTAAAGACGCTTCTGTTAGCGCG 300  
Db 11441 CATCAACCTGGTATCATGATCATCTCCGGGTAAAGACGCTTCTGTTAGCGCG 11382  
QY 301 TGAACCTGGCAGAGCGAATGTTTGGCTGATGTTCTGACCTGGCGGAGCAAGT 360  
Db 11381 TGAACCTGGCAGAGCGAATGTTTGGCTGATGTTCTGACCTGGCGGAGCAAGT 11322  
QY 361 CGATAAATTCCTGGCGCTCGAATTCGTCAGATGATCATCAACCAACCGTTCAACC 420  
Db 11321 CGATAAATTCCTGGCGCTCGAATTCGTCAGATGATCATCAACCAACCGTTCAACC 11262  
QY 421 GCGTGAACGATTCGTGACGCAACCTAAGTCCGCTACCAATCTGGGTACTGT 480

Db 11261 GCGTGAACGACGATTCGTGCACGCAACCTGCTCCCGACATGATCTGGGTACTGT 11202  
Qy 481 CAGCGAAGACGTCGTAGCGTTGAAAGCTACAGTTCAATGTTGGAACTGGACATCA 540  
Db 11201 CAGCGAAGACGTCGTAGCGTTGAAAGCTACAGTTCAATGTTGGAACTGGACATCA 11142  
Qy 541 CAGCGCTTGTGATGCGCCCTGATGCGAGCACTACAGACTCCGCGCAGCGAGTTCCG 600  
Db 11141 CAGCGCTTGTGATGCGCCCTGATGCGAGCACTACAGACTCCGCGCAGCGAGTTCCG 11082  
Qy 601 CCGCATGCTTCACTTCTGTGAAAAACCGCAAAATTCAGTCCCGTGTAACTGCTGAA 660  
Db 11081 CCGCATGCTTCACTTCTGTGAAAAACCGCAAAATTCAGTCTGTGCTGAACTGCTGAA 11022  
Qy 661 GAAATAGACCGCGCGTAGCTGAAACCGCAGCAGCCGACTGTAGAGTGCATGCGTCCGCG 720  
Db 11021 GAAATAGACCGCGCGTAGCTGAAACCGCAGCAGCCGACTGTAGAGTGCATGCGTCCG 10962  
Qy 721 TATTCGTAAACATTTGCAATCTACGCGGATACCGCGGAATCATGCGCACATTCACGG 780  
Db 10961 TATTCGTAAACATTTGCAATCTACGCGGATACCGCGGAATCATGCGCACATTCACGG 10902  
Qy 781 TGAAGTTATCGCTTCTGCGGTATCTGGAAGTTAATCGGCTTTACACCGTCAAAAA 840  
Db 10901 TGAAGTTATCGCTTCTGCGGTATCTGGAAGTTAATCGGCTTTACACCGTCAAAAA 10843  
Qy 841 AACGGGCTTTAGGCGGCTTTTATTTTCAACCTATTTCAGATAGTAAGTATC 900  
Db 10842 AACGGGCTTTAGGCGGCTTTTATTTTCAACCTATTTCAGATAGTAAGTATC 10783  
Qy 901 GTCCGTTGTAACCTCTTTACTGCGCTT 927  
Db 10782 GTCCGTTGTAACCTCTTTACTGCGCTT 10756

RESULT 9  
LOCUS AE016993/c 225944 bp DNA linear BCT 22-APR-2003  
DEFINITION Shigella flexneri 2a str. 2457T section 16 of 16 of the complete  
ACCESSION AE016993 AE014073  
VERSION AE016993.1 GI:30043695  
KEYWORDS  
SOURCE  
ORGANISM Shigella flexneri 2a str. 2457T  
REFERENCE 1 Bacteri; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Shigella.  
1 (bases 1 to 225944)  
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,  
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,  
Schwartz,D.C. and Blattner,F.R.  
Complete Genome Sequence and Comparative Genomics of Shigella  
flexneri Serotype 2a Strain 2457T  
Infect. Immun. 71 (5), 2775-2786 (2003)  
12704152  
TITLE  
JOURNAL 2 (bases 1 to 225944)  
PUBMED Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,  
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,  
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,  
Schwartz,D.C. and Blattner,F.R.  
Direct Submission  
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source  
1. 225944  
/organism="Shigella flexneri 2a str. 2457T"  
/mol\_type="genomic DNA"  
/strain="2457T"  
/serotype="2a"  
/db\_xref="taxon:198215"  
gene complement (271..867)

CDS  
/gene="fimb"  
/locus\_tag="S4466"  
complement (271..867)  
/gene="fimb"  
/locus\_tag="S4466"  
/function="regulator; Surface structures"  
/note="residues 1 to 198 of 198 are 100.00 pct identical  
to residues 1 to 198 of 198 from Escherichia coli K-12 :  
B4313; involved in phase variation"  
/codon\_start=1  
/transl\_table=11  
/product="recombinase; regulator for fima"  
/protein\_id="AAP19415.1"  
/db\_xref="GI:30043696"  
/translation="MSKRRYLTCGEVQAMQVCYATGARDYCLILAYRHGRISB  
LIDLIYODDLNENRINIRLNKGFSTVPLRDEAEVETPOEERANMGARDAI  
FISRRSPRLSROOAVRIIDAGIEACTVOTPHMRHCGYELARQADTRILQYLV  
GHRNIRHTVRYTNSNARFAGLMERNLNEKLRREV"  
complement (1462..1947)  
/gene="fimb"  
/locus\_tag="S4467"  
complement (1462..1947)  
/gene="fimb"  
/locus\_tag="S4467"  
/function="regulator; Surface structures"  
/note="residues 1 to 160 of 199 are 97.50 pct identical to  
residues 1 to 160 of 200 from Escherichia coli K-12 :  
B4312; involved in phase variation"  
/codon\_start=1  
/transl\_table=11  
/product="recombinase; regulator for fima"  
/protein\_id="AAP19416.1"  
/db\_xref="GI:30043697"  
/translation="MRNKAQNKKNFLTSEISLLKAANTGPHATRYCLILCFIH  
GRASRICKLRISDIDLKAKCIYIRHLKKGFSSTHRLKVEQALKNWLSIRPSVHA  
ESBWFLSRKNPLSRQPFHIIISGNAJSLERHPMLRISCFALANMGIDTRL  
I"

gene  
3569..4117  
/gene="yJha"  
/locus\_tag="S4468"  
3569..4117  
/gene="yJha"  
/locus\_tag="S4468"  
/note="residues 1 to 182 of 182 are 100.00 pct identical  
to residues 60 to 241 of 241 from Escherichia coli K-12 :  
B4311"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP19417.1"  
/db\_xref="GI:30043698"  
/translation="MESNTWTIHNKKENALNDQVEVNYAIKLDQDQTVRPGMLT  
HFSNGTRIGPYVKSMDATKDLNFRIRRYRKAYARQDLSGDSNRDNRHRRDYVT  
YHNSDFTFAMQDTTLYSKQNDRYRANHKWATENAFVLDYHMTPTDTPYEYDLRO  
GVYGRNDLSNSYRIGVSFKL"  
4029..5243  
/gene="yJht"  
/locus\_tag="S4469"  
4029..5243  
/gene="yJht"  
/locus\_tag="S4469"  
/note="residues 1 to 404 of 404 are 98.76 pct identical to  
residues 1 to 404 of 404 from Escherichia coli K-12 :  
B4310"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAP19418.1"  
/db\_xref="GI:30043699"  
/translation="MTTLTARVPTTAEIIRYRKTVIALVNLCSROEVTYTNKTIATL  
ATIMASFAANASVLPETVPYPSGTAINDVYIGLSAGATWYKLDLPQADKKRTA  
LAAFPGGPREQATISAFIDNLVYFGIGNSGLTQVDFVHKYINKTKNSLASHA  
PMGMGHVTFVHNKRAYVTGVNOMIFNGYFEDLNEAGDSTALDKINAHYEDKQAD





```
misc_feature      complement(3375..3404)
                  /gene="STY4861"
                  /note="PS00120 Lipases"
gene              4066..4839
                  /gene="uxur"
                  /note="synonym: STY4862"
CDS               4066..4839
                  /gene="uxur"
                  /note="Similar to Escherichia coli uxu operon
transcriptional regulator uxur SW:UXUR_ECOLI (939161) (257
aa) fasta scores: E(): 0, 89.9% id in 257 aa, and to
Brwinia chrysanthemi exu regulon transcriptional regulator
exu SW:EXUR_ERWCH (Q9X9E0) (259 aa) fasta scores: E(): 0,
46.9% id in 256 aa. Contains helix-turn-helix motif,
residues 34 to 55, score 975(+2.51 SD).
Fasta hit to PDHR_ECOLI (254 aa), 32% identity in 221 aa
Overlap
Fasta hit to EXUR_ECOLI (258 aa), 48% identity in 255 aa
Overlap
Orthologue of E. coli uxur (UXUR_ECOLI); Fasta hit to
UXUR_ECOLI (257 aa), 90% identity in 257 aa overlap"
/codon_start=1
/transl_table=1
/product="uxu operon transcriptional regulator"
/protein_id="CAD03351.1"
/db_xref="GI:16505623"
/db_xref="GOA:Q820Y1"
/db_xref="TrEMBL:Q820Y1"
/translacion="MKSNTSQRPYQEVGAMIRDLIVQPPRGERLPPERIAERLN
VTRIVVEALIMLEIKGLVEVRGAGIYVDSADNNMEGADYVHCHNDGAPPELLAR
QLLESNIAEFALQATREDIIRKQALQLEERLASAPRGSSGMORHIAEYTH
NSMIVELFRSQMQRNNPMWLDHSHLGTITRKMLGDHKOILALIKKQARAATL
AMWQHELVYKRLRPSNVDDITFDGILFESWPLDVDA"
4105..4284
/gene="uxur"
/note="Pfam match to entry PF00392 gntR, Bacterial
regulatory proteins, gntR family, score 107.20, E-value
1.4e-31"
4162..4236
/gene="uxur"
/note="PS00043 Bacterial regulatory proteins, gntR family
signature"
complement(4840..5853)
/gene="trpS2"
/note="synonym: STY4863"
complement(4840..5853)
/gene="trpS2"
/EC_number="6.1.1.2"
/note="Similar to Clostridium longisporum
tryptophanyl-tRNA synthetase Trps or Trsa SW:SYW_CLOLO
(Q46127) (341 aa) fasta scores: E(): 0, 54.8% id in 330
aa, and to Thermotoga maritima tryptophanyl-tRNA
synthetase Trps SW:SYW_THEMA (Q9WY2) (328 aa) fasta
scores: E(): 0, 36.5% id in 329 aa. Note that the E.coli
trps ortholog is STY4315"
/codon_start=1
/transl_table=1
/product="probable tryptophanyl-tRNA synthetase"
/protein_id="CAD03352.1"
/db_xref="GI:16505624"
/db_xref="GOA:Q820Y0"
/db_xref="TrEMBL:Q820Y0"
/translacion="NNTLPILITGDRPTGSLHLGHVSLRORVALOODHOQVLIAD
LOGLTGNSNPQIKRDNIPQVADYLAVIDPALTTCISALPRLATLVYNNITV
VARVERNPVKNKELQKGFRTSLPYGFAPYPISSQADITTKAKENVPGDQJLPMIQ
TNEIVHKNNLSFSSPYLPPCCALSDTRLELFDSSAVMSLVNTLLLSGSEETIHR
AVSAMYPYNNHKISDPKIRGNVFTWLDLPHDPKAVAMKMYQCGGSGDVCCNK
ELFETLQELIAPIRRRATFIADKGMIMELIKGSERAHEVTOKTLQEVKKGILPLTL
FQV"
complement(5170..5835)
/gene="trpS2"
/note="Pfam match to entry PF00579 trna-synt_1b, trna
```

```
Query Match      77.2%; Score 715.6; DB 1; Length 88037;
Best Local Similarity 92.2%; Pred. No. 2.9e-201;
Matches 754; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY      1  GTCATGTTACGGCGATCATGTTAATTTGACAGATCATGACGAGTTCAGGACTTTTGT 60
DB      86791 GTCATGTTACGGCGATCATGTTAATTTGACAGATCATGACGAGTTCAGGACTTTTGT 86732

QY      61  ACTTCCTGTTTCGATTTAGTTCGCAATTTAGGTAGCAACATGACAGCCGCCACATTTCT 120
DB      86731 ACTTCCTGTTTCGATTTAGTTCGCAATTTAGGTAGCAACATGACAGCCGCCACATTTCT 86672

QY      121 TATCGTTGAAGCGAGTTGGTGTAAACGCAACGTTGAAAAGTATTTTGAAGCGAAGC 180
DB      86671 TATCGTTGAAGCGAGTTGGTGTAAACGCAACGTTGAAAAGTATTTTGAAGCGAAGC 86612

QY      181 CTATGATGTTTTGAAAGCAGACAGATGCGCGGAAATGCAATCGATCTCTCGAATATGA 240
DB      86611 CTATGATGTTATCGAAGCAGACAGATGCGCGGAAATGCAATCGATCTCTCGAATATGA 86552

QY      241 CATCAACCTGGTGATCATGATATCATATCTCGCGGTAAAGACGGCTCTTGTTAGCGCG 300
DB      86551 CATCAACCTGGTGATCATGATATCATATCTCGCGGTAAAGACGGCTCTTGTTAGCGCG 86492

QY      301 TGAACCTGCGCGAGCGAGGAAATGTTGCGGTATGTTCTGACTGCGCGGTGCAACGAAGT 360
DB      86491 TGAACCTGCGGTAGAGCGCTAATGTTGCGGTATGTTCTGACTGCGCGGTGCAACGAAGT 86432

QY      361 CGATTAATTTCTCGGCTCGAAATGCGTGTGACATGATCATCAACCAACCGTTCAACCC 420
DB      86431 CGACCAAAATTTCTCGGCTCGAAATGCGGCGGTGATGATCATCAACCAACCGTTTAAATCC 86372

QY      421 GGTGACCTGACGATTCGTGACGCAACCTACTCTGCTCCGTACATGAATCTGGTACTGT 480
DB      86371 AGCGAATTTGACATTCGTGCGCCGTAACCTGCTCTCCGTACATGAACCTGGGTACTCGT 86312

QY      481 CAGCGAAGACGTGTAAGCTTTGAAAGCTTCAAGTTCAATGTTGGGAACTGACATCAA 540
DB      86311 CAGCGAAGACGTGTAAGCTTTGAAAGCTTCAAGTTCAATGTTGGGAACTGACATCAA 86252

QY      541 CAGCGTTTCGTTGATTCGCGCTGTATGCGAGACAGTCAAGCTGCGCGCGAGCGATTCCG 600
DB      86251 CAGCGTTTCGTTGATTCGCGCTGTATGCGCGAGACAGTCAAGCTGCGCGCGAGCGATTCCG 86192

QY      601 CGCATGCTTCACTTCTGTGTAACCCAGGCAAAATTCAGTCCCGTCTAACTGCTGAA 660
DB      86191 CGCATGCTTCACTTCTGTGTAACCCAGGCAAAATTCAGTCCCGTCTGAGCTGCTGAA 86132

QY      661 GAAATGACCGGCGGTAGCTGAAACCGACAGACCGTACTGTAGACGTGACGATCCGCGC 720
DB      86131 GAAATGACCGGCGGTAGCTGAAACCGACAGACCGTACTGTAGACGTGACGATCCGCGC 86072

QY      721 TATTCGTAACATTTGCAATCTACCGCGGTATACCGCGGAAATCATATGCGCCACATTCACGG 780
DB      86071 TATTCGTAACATTTGCAATCTACCGCGGTATACCGCGGAAATCATATGCGCCACATTCACGG 86012

QY      781 TGAAGTTATCGCTTCTGCGGTGATCGAAGATTAT 818
DB      86011 CGAAGTTATCGCTTCTGCGGTGATCGAAGATTAT 85974

RESULT 11
AB016849/c      28423 bp      DNA      linear      BCT 21-MAR-2003
LOCUS           Salmonella enterica subsp. enterica serovar Typhi Ty2, section 16
DEFINITION      AB016849 AB014613
ACCESSION       AB016849
VERSION         AB016849.1 GI:29140243
KEYWORDS
SOURCE          Salmonella enterica subsp. enterica serovar Typhi Ty2
ORGANISM        Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
```



```
REFERENCE      1 (bases 1 to 284233)
AUTHORS        Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
                Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
TITLE          Comparative Genomics of Salmonella enterica Serovar Typhi Strains
                Ty2 and CT18
JOURNAL        J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE        2 (bases 1 to 284233)
                12644504
                PubMed
PUBMED        12644504
AUTHORS        Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
                Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.
TITLE          Direct Substitution
                Submitted (25-SEP-2002) Laboratory of Genetics, University of
                Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES       location/Qualifiers
                1..284233
                /organism="Salmonella enterica subsp. enterica serovar
                Typhi Ty2"
                /mol_type="genomic DNA"
                /strain="Ty2"
                /serovar="Typhi"
                /sub_species="enterica"
                /db_xref="taxon:209261"
                301..489
                /locus_tag="t4354"
                /locus_tag="t4354"
                301..489
                /locus_tag="t4354"
                /locus_tag="t4354"
                /note="corresponds to STY4663 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071807.1"
                /db_xref="GI:29140244"
                /translation="MTGKPSERTGTGRTIGSEMMVRCDFGNEYLIHAGEAFVSENHDA
                WVVGDTPCVALDTHFLR"
                614..2125
                /locus_tag="t4355"
                614..2125
                /locus_tag="t4355"
                /locus_tag="t4355"
                /note="corresponds to STY4664 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative DNA helicase"
                /protein_id="AA071808.1"
                /db_xref="GI:29140245"
                /translation="MPHTDPEQAAVIANKGERLVVCAFGSGKTTTLRRVAENPTE
                RMUYIAVNRARIDAEQKFPFNTCTSHQLAWPTVGKYSRLINSLTYVARALN
                SRNMLARLTLDVLRNFICSAATDITPEHNPDDCKGLOPQIILPSAKIMAMSGAR
                ODGPEVTHDTYVLTLYOLSKPNLSRYTTLIPDAADANPVISAIVLNONCRVVLVGD
                HQIYRFRGADNANMOPQLENSDRMLVTHSFRGPEVAGVANRLMLAKGETHOITKGG
                GAGRVLLTWRNFGHVAHRSVCGVIRTAHLMSLGRVFPWVGWESYRIEDLDLY
                WSDIMSEKRRHRLTRYRDYDEYLIADETGDVMKOGIFLEQFPPLPRNLTLR
                BORVISEADYVCTAHSKSGLEMDRVKLHDFDLIDPDMESQROBINIMVYAA
                TSARKVLTDPILTHLLIQPRADIAGTSADSKSGSESEVSTENDKGTETVYVPS"
                2109..3698
                /locus_tag="t4356"
                2109..3698
                /locus_tag="t4356"
                /note="corresponds to STY4665 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071809.1"
                /db_xref="GI:29140246"
                /translation="MRKYKSMWEKKNFROIENKSGNAKKRLIPTDGYFNPALISD
                LIQHERROOMRLILMENSALPKRPERFYLLPHGVVGCORLPASAGKFPATDQW
                DVIOITTVFAVRLSKGYMLPRGASAEQASQSVAVVYVYALPHSLGRNINTEBL
                RGAVWRPGLSVPEEPYRFRKAEPIAGQVIGELIALRFLPEPVLOWKGNPDLR
                TLAPISGRYSADMIDTIDVINEAIVHAGGIPLGFPVIAEGEQTSELAELINPAVAVPV
                LKKSSTTPPDVAVMWVGQETTVLLSLSDSLPSDAKPKRGQPTAETDPDVLQVMSITLG
                GLTNTDTQSGPEKEGVNTEITDMVAGSEFPEVAVALFETSAABEVINDVTTECTGSSGCG
                ISLSSTELGDLFWMSWLRDGLREGDIPVNTADACVHLTCGFVPI SVPGVFFLKSHS
                RSCSSGSLKSGRKEQVOAAPEKMRKRVSDSRFMQCCVIEEPGGRGVKLTGYLIK
                MSEIYANGNPPDLSFLKVIN"
                3662..4875
                /locus_tag="t4357"
                3662..4875
                /locus_tag="t4357"
                /note="corresponds to STY4666 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="probable phage integrase"
                /protein_id="AA071810.1"
                /db_xref="GI:29140247"
                /translation="METNTTMOOLIDEYFPAPKRLRSASESYTKVFPVWGPLSC
                PNDVTHKYLARRRLTKKKSAGRTNNKVAHMRALPNYGIORGILHLDENFPNNSV
                VEPDKRKKTLLTOAQLEIVAYQIMEQYENENGTGLKYSRCALEPMFWLTVLDTLY
                TGIROMQLHRLNDVDLREGQIRLITTECKKHKEYVPIISFLRPLTCLMEKAQSE
                GLKGNDRLEFNIALFTGKDPALIGDMDSPOVRAVFRRLSECCFALISPHFRPLATEM
                MKKPEQNLHMAQSVLGHSMNKSTLEVENDIAMVGRALBAQFMQIKAAVARSISGLT
                KRR"
                5302..5595
                /locus_tag="t4358"
                5302..5595
                /locus_tag="t4358"
                /note="corresponds to STY4667 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="AA071811.1"
                /db_xref="GI:29140248"
                /translation="MPVANSMAKRETNLRIRKPARBDLIDRAAKARGKNCIDFVLEA
                AAAAEBALIBORIIMADPEAYQAFILRLOAPAPADALRKTWQTPAPWQER"
                5592..6080
                /locus_tag="t4359"
                5592..6080
                /locus_tag="t4359"
                /note="corresponds to STY4668 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="conserved hypothetical protein"
                /protein_id="AA071812.1"
                /db_xref="GI:29140249"
                /translation="MISAPEPLHAGHILTPFCCGDSMDNMLKORAMKQVTCASRTF
                VCCGSDVLAAYSLASSAVTNTAPGRRRAMPDPIPVVILGRILAVDRSLHQGVGR
                ALVRDAGRLVIOVAETIGIRGMLVHALSDAREFYLVRGVFEPSPMDPMLMVLGDLV
                HA"
                complement(6260..6712)
                /locus_tag="t4360"
                complement(6260..6712)
                /locus_tag="t4360"
                /note="corresponds to STY4669 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="putative membrane protein"
                /protein_id="AA071813.1"
                /db_xref="GI:29140250"
                /translation="MEQCKNDAILIEHKYNSKHIDFERSQANSQIGWLFI STLGCWSV
                NPLIOVATAIILLFCIFPNSKODMTERKAPFKRIEDIAKDDISNLIGSRRAKRLVDL
                GLVEKTKRAIKVPLKTSPIFTVICYIFYSISFLVPSNLPFRKRLIFNP"
                complement(7349..9634)
                /locus_tag="t4361"
                complement(7349..9634)
                /locus_tag="t4361"
                /note="corresponds to STY4670 from Accession AL513382:
                Salmonella typhi CT18"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein"
```





```
/gene="sthd"
/Note="similar to E. coli fibrial morphology
(AAC7275.1); Blastp hit to AAC7275.1 (167 aa), 31%
identity in aa 18 - 166"
/codon_start=1
/transl_table=11
/product="putative fibrial subunit"
/protein_id="AAL23407.1"
/db_xref="GI:16423159"
/translation="MMNRFPALALITFTGLAMSPSISAAYNTADAVGINTGLIY
ADPCIVTPDTLSIGQYARADJAVAGNRTATPFTIKMTCATTTKAIIFNVPYR
DPYASAIYANSANPAODLQLEFMDKSLVLANVYTVTDINSNTBANLPAVA
RMVTPGKVTAGDFISVITITAMQ"
complement(5924..5929)
/gene="sthd"
/Note="putative RBS for sthd; RegulonDB:STM51H004452"
```

## RBS

```
Query Match 77.0%; Score 714; DB 1; Length 13852;
Best Local Similarity 92.1%; Pred. No. 6,1e-201;
Matches 753; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 GTGATGTTACCGCGATCATGTTAATTGAGCATCATCAGCAGGTGAGGACTTTTGT 60
DB 12606 GTGATGTTACCGCGATCATGTTAATTGAGCATCATCAGCAGGTGAGGACTTTTGT 12547
QY 61 ACTTCCTGTTGCAATTGAGTGGCAATTAGTAGCAACATGCAAGACCCGCAATTCT 120
DB 12546 ACTTCCTGTTGCAATTGAGTGGCAATTAGTAGCAACATGCAAGACCCGCAATTCT 12487
QY 121 TATCGTTGAAGCAGGTTGTTAACAACGCAACAGTTGAAAAGTATTTTTCGAACGGGAAG 180
DB 12486 TATCGTTGAAGCAGGTTGTTAACAACGCAACAGTTGAAAAGTATTTTTCGAACGGGAAG 12427
QY 181 CTATGATGTTTTCGAAGCAGACAGATGGCGGGAATGCATGATCCTCTCTGAATATGA 240
DB 12426 CTATGATGTTTTCGAAGCAGACAGATGGCGGGAATGCATGATCCTCTCTGAATATGA 12367
QY 241 CATCAACCTGGTATCATGATATCATCTCCGGGTGAGAACAGTCTTCTGTTACGGG 300
DB 12366 CATCAACCTGGTATCATGATATCATCTCCGGGTGAGAACAGTCTTCTGTTACGGG 12307
QY 301 TGAACGTGGGAGAGGCGAATGTTGGTGTGATGTTCTGTAATGCGCGTGAACAAGAGT 360
DB 12306 TGAACGTGGGAGAGGCGAATGTTGGTGTGATGTTCTGTAATGCGCGTGAACAAGAGT 12247
QY 361 CGATTAATTTCTGGGCTCGAATTCGGTGAAGTGAATGATCATCAACAAACGTTTAAACC 420
DB 12246 CGATTAATTTCTGGGCTCGAATTCGGTGAAGTGAATGATCATCAACAAACGTTTAAACC 12187
QY 421 GCGTGAACGTAGATTCGTGCAAGCACTGCTGCCGTACATGAATCTGGGACTGT 480
DB 12186 GCGTGAACGTAGATTCGTGCAAGCACTGCTGCCGTACATGAATCTGGGACTGT 12127
QY 481 CAGCGAAGAACGTGTGAGGTTGAAAGCTACAAGTTCAATGTTGGGAACTGGACATTA 540
DB 12126 CAGCGAAGAACGTGTGAGGTTGAAAGCTACAAGTTCAATGTTGGGAACTGGACATTA 12067
QY 541 CAGCGTGTGATGATGGCCCTGATGGGAGAGTAAAGCTGCGCGCAAGATTCCG 600
DB 12066 CAGCGTGTGATGATGGCCCTGATGGGAGAGTAAAGCTGCGCGCAAGATTCCG 12007
QY 601 CCGCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGTGAACGTCTGAA 660
DB 12006 CCGCATGCTTCACTTCTGTGAAAACCCAGGCAAAATTCAGTCCGTGTGAACGTCTGAA 11947
QY 661 GAAATTCAGCCGCGCTGAGCTGAAAACCGCAGACCGTATGTAAGCTGACGATCCGCG 720
DB 11946 GAAATTCAGCCGCGCTGAGCTGAAAACCGCAGACCGTATGTAAGCTGACGATCCGCG 11897
QY 721 TATGTGTAACATTTGCAATCTAAGCCGAGATACCCGGAATCATGGCCACATTCACG 780
DB 11896 TATGTGTAACATTTGCAATCTAAGCCGAGATACCCGGAATCATGGCCACATTCACG 11827
```

```
QY 781 TGAAGTTATCGCTTCTGGCGGTATCTGGAAGATTAT 818
DB 11826 CGAAGTTATCGCTTCTGGCGGTATCTGGAAGATTAT 11789

RESULT 13
AR385083
LOCUS AR385083 726 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 1812 from patent US 6610836.
ACCESSION AR385083
VERSION AR385083.1 GI:40094817
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 726)
AUTHORS Breton G.L. and Osborne M.
TITLE Nucleic acid amino acid sequences relating to Klebsiella pneumoniae
JOURNAL for diagnostics and therapeutics
PATENT: US 6610836-A 1812 26-AUG-2003;
FEATURES
source
location/Qualifiers
1..726
/organism="Unknown"
/mol_type="genomic DNA"
ORIGIN
```

```
Query Match 62.3%; Score 577.2; DB 6; Length 726;
Best Local Similarity 87.2%; Pred. No. 2,1e-160;
Matches 633; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

```
QY 92 GTAGCAAAACATGCAAGACCCGCACTTTCTTATGTTGAAGACAGTGTGTAACACGCAAC 151
DB 1 GTAGCAAAACATGCAAGACCCGCACTTTCTTATGTTGAAGACAGTGTGTAACACGCAAC 60
QY 152 AGTTGAAAAGTATTTTCGAAGCGAAGGCTATGATGTTTGAAGCGCAATGTCGCG 211
DB 61 AGTTGAAAAGTATTTTCGAAGCGAAGGCTATGATGTTTGAAGCGCAATGTCGCG 120
QY 212 GAATGATGATGATCTCTCTGTAATATGATCAATCAACCTGGTATCATGATATCAATCTG 271
DB 121 GAATGATGATGATCTCTCTGTAATATGATCAATCAACCTGGTATCATGATATCAACCTG 180
QY 272 CCGGGTGAAGACCGTCTTCTGTTAGCGCGTGAACCTGCGGACGCAATGTTGCGTGG 331
DB 181 CCGGGTGAAGACCGTCTTCTGTTAGCGCGTGAACCTGCGGACGCAATGTTGCGTGG 240
QY 332 ATGTTGCTGATCGCGCGTGAACAAGATGATTAATTTCTGCGCTTGAATCGTGCA 391
DB 241 ATGTTGCTGATCGCGCGTGAACAAGATGATTAATTTCTGCGCTTGAATCGTGCA 300
QY 392 GATGACTATCATCAACCAACCGTTCAACCGGCTGAACAGATTCGTGACGCACTTA 451
DB 301 GATGACTATCATCAACCAACCGTTTAAACCGGCTGAACAGATTCGTGACGCACTTA 360
QY 452 CTGTCCCGTATCATTAATTCGGTATCTGTCAAGCAAGACGTGTGACGTTGAAGCTAC 511
DB 361 CTGTCCCGTATCATTAATTCGGTATCTGTCAAGCAAGACGTGTGACGTTGAAGCTAC 420
QY 512 AAGTTCAATGTTGGGAACTGACATCAACAGCCGTTGATCGGCTGTGATGGCGAG 571
DB 421 AAGTTCAATGTTGGGAACTGACATCAACAGCCGTTGATCGGCTGTGATGGCGAG 480
QY 572 CAGTTCAGAGCTGCGCGAGGAGTTCGCGGCAATGCTTCACTTCTGTGAAAACCGAGCG 631
DB 481 CAGTTCAGAGCTGCGCGAGGAGTTCGCGGCAATGCTTCACTTCTGTGAAAACCGAGCG 540
QY 632 AAAATTGAGTCCGCTGTGAACTGTGAAAGAAAATGAACGCGCGTGAAGCTGAAACGCGAC 691
DB 541 AAAATTGAGTCCGCTGTGAACTGTGAAAGAAAATGAACGCGCGTGAAGCTGAAACGCGAT 600
QY 692 GACCGTATCTGTGAAGCAGATCCGCGTATTCGTTAACTTTGGAATCTTACGCGCGAT 751
DB 601 GACCGTATCTGTGAAGCAGATCCGCGTATTCGTTAACTTTGGAATCTTACGCGCGAT 660
```

QY	752	ACGGCGAAATCATGCGACCACTTGAAGGTATTCCTCTGCGGATCTGGAA	811
Db	661	ACCCGGAATCATCTCAATCACTACCGATGAAGCTACCGTTCTGTGGGATCTGCAG	720
QY	812	GATTAA	817
Db	721	GAAATTA	726
RESULT 14			
AF267619	564 bp	DNA	linear
LOCUS	Escherichia coli strain E2348/69	aerobic regulator (arcA) gene,	
DEFINITION	partial cds.		
ACCESSION	AF267619	GI:9664513	
VERSION	AF267619.1		
KEYWORDS			
SOURCE	Escherichia coli		
ORGANISM	Escherichia coli		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
AUTHORS	1 (bases 1 to 564) Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and Whitlam,T.S.		
TITLE	Parallel evolution of virulence in pathogenic Escherichia coli		
JOURNAL	Nature 406 (6791), 64-67 (2000)		
MEDLINE	20351039		
PUBMED	10894541		
REFERENCE	2 (bases 1 to 564) Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and Whitlam,T.S.		
AUTHORS	Submitted (12-MAY-2000) Biology, Pennsylvania State University, 208 Mueller Laboratory, University Park, PA 16802, USA		
TITLE	Direct Submission		
JOURNAL	Location/Qualifiers		
FEATURES	1..564		
SOURCE	/organism="Escherichia coli" /mol_type="genomic DNA" /strain="E2348/69" /serotype="O127:H6" /specimen_voucher="Taunton 69" /db_xref="taxon:562" /note="EPEC 1" <1..>564 /gene="arcA" <1..>564 /gene="arcA" /codon_start=1 /transl_table=1 /product="aerobic regulator" /protein_id="AAF97155.1" /db_xref="GI:9664514" /translation="ATDGAENHQILSEYDINLVIMDINLPKGNLLIARBLREQANVALMPFTGRDNEVDKILGLEIGADYITKFPNRELTIRANILSTRNMLGTVSEBRRSVLSYKNGWELDINSRLIGPDGEQYKLPKRSFRAMLPKCNPKIGSRBELKKMTGR ELKPHDRIVDVITIRIKHFESTDPE"		
ORIGIN			
Query Match	60.7%	Score 562.4	DB 1; Length 564;
Best Local Similarity	99.8%	Pred. No. 5.2e-156;	
Matches 563;	Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
QY	197	GCGACAGATGCGCGCGGAAATGATCATCAATCTCTTGTAATATGACATCAACCTGATGTC	256
Db	1	GCGACAGATGCGCGCGGAAATGATCATCAATCTCTTGTAATATGACATCAACCTGATGTC	60
QY	257	ATGATATCATCTGCGCGGTAAGAACGCTCTTCTGTAGGCGCGTGAACCTGCGAGAG	316
Db	61	ATGATATCATCTGCGCGGTAAGAACGCTCTTCTGTAGGCGCGTGAACCTGCGAGAG	120
QY	317	GCGAATGTTGCGTGTGATTTCTTGAATGCGCGTGAACGAAATGATTAATTTCTCGGC	376

D	b			GCGAATGTTGGCTTGAATGTTTCCTGACTCAGGCCGTGCACAACGAAGTGCAATAAAATTCTCGGC	180
Oy		121			
Oy		377		CTCGAATACTGGTGACATGACTCATCATCCAAACCGTTCAAACCCGCGTGAACGTACGATTT	436
D	b			CTTCGAAAATCGGTGACGATGACTCATCATCAACAAACCGTTCAACCCGCGTGAACGTACGATTT	240
Oy		437		CCTCACAGCAACCTAACGTCCTCCCCTGATCCATGTAATCTGGGGTACTGTACAGCAAGAAGCGTCG	496
D	b			CGTCACAGCAACCCTGCTGTCCTCCCTGATCCATGTAATCTGGGGTACTGTACAGCAAGAAGCGTCG	300
Oy		497		AGCGTTGAAAGCTCAAGTTCAATGTTGGGAATGACATCAACACCCGTTGGTAGTC	556
D	b			301 AGCGTTGAAAGCTCAAGTTCAATGTTGGGAATGACATCAACACCCGTTGGTAGTC	360
Oy		557		GGCCCTGATGCGAGCAAGTACMAAGCTCCCGCGACGAGATTCCGCGCATGCTTCACTTC	616
D	b			361 GGCCCTGATGCGAGCAAGTACMAAGCTCCCGCGACGAGATTCCGCGCATGCTTCACTTC	420
Oy		617		TGTGAAAAACCGACCAAAATTCAGTCCCCCTGCTGAATCTGCTGAAGAAAATGACCGGCGCT	676
D	b			421 TGTGAAAAACCGACCAAAATTCAGTCCCCCTGCTGAATCTGCTGAAGAAAATGACCGGCGCT	480
Oy		677		GAGCTGAAAACCGACGACCGGTAAGTGAAGTGAAGTCCGCGCATGCTTCACTTC	736
D	b			481 GAGCTGAAAACCGACGACCGGTAAGTGAAGTGAAGTCCGCGCATGCTTCACTTC	540
Oy		737		GAATCTAACCGCGGATACGCGCGAA	760
D	b			541 GAATCTAACCGCGGATACGCGCGAA	564
RESULT_15					
LOCUS		AF267620		564 bp DNA linear BCT 03-AUG-2000	
DEFINITION		Escherichia coli strain DECla aerobic regulator (arca) gene,			
ACCESSION		AF267620			
VERSION		AF267620.1		GI:9664515	
KEYWORDS					
SOURCE					
ORGANISM				Escherichia coli Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	
REFERENCE		AUTHORS		Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and Whittem,T.S. Parallel evolution of virulence in pathogenic Escherichia coli	
JOURNAL MEDLINE PUBMED		PUBLISHED		Nature 406 [6791], 64-67 (2000)	
REFERENCE		AUTHORS		Reid,S.D., Herbelin,C.J., Bumbaugh,A.C., Selander,R.K. and Whittem,T.S. Direct Submission Submitted (12-MAY-2000) Biology, Pennsylvania State University, 208 Mueller Laboratory, University Park, PA 16802, USA	
FEATURES		SOURCE		location/Qualifiers	
		gene		1..564 /organism="Escherichia coli"	
		CDS		/mol_type="genomic DNA" /strain="DECla" /serotype="O5:H6" /specimen_voucher="CDC 572-56" /db_xref="Taxon:562" /note="EPEC 1"	
				<1..>564 /gene="arca"	
				<1..>564 /gene="arca"	
				/codon_start=1 /transl_table=11 /product="aerobic regulator" /protein_id="AAF97160.1"	

/db\_xref="GI:9664516"  
/translation="ATDGAEMHQILSEVDINLYIMDINLPGRNGLILARELREQANVA  
LMFLTGDRDNEVDKILGLEIGADDYITKPEVNPRLTIRARNLISRTNLCITVSEERSV  
ESYKFNQMEHDINSRLIGPDGEYKLPSPSEPRAMLHPCENPQKIQSRRELKMKMTGR  
ELKPHDRTVDVITIRIRIKHESTPDPTE"

## ORIGIN

```
Query Match      60.7%; Score 562.4; DB 1; Length 564;  
Best Local Similarity 99.8%; Pred. No. 5.2e-156;  
Matches 563; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY      197 GCGACAGATGCGCGGAAATGATCATGATCCTCTGAAATATGATCATCACTGGTGATC  
DB      1 GCGACAGATGCGCGGAAATGATCATGATCCTCTGAAATATGATCATCACTGGTGATC 60  
QY      257 ATGATATCAATTCGCGGTAAGAACGGTCTTCTTTAGCCGTGAACCTGCGAGCAG 316  
DB      61 ATGATATCAATTCGCGGTAAGAACGGTCTTCTTTAGCCGTGAACCTGCGAGCAG 120  
QY      317 GCGAATGTTGCGTTGATGTTCCGACTGCGCCGTGACAAAGATGATTAATTCGCGC 376  
DB      121 GCGAATGTTGCGTTGATGTTCCGACTGCGCCGTGACAAAGATGATTAATTCGCGC 180  
QY      377 CTGGAATGTTGTCAGATGACTACATCAACAAACCGTCAACCGCGTGAACGATC 436  
DB      181 CTGGAATGTTGTCAGATGACTACATCAACAAACCGTCAACCGCGTGAACGATC 240  
QY      437 CGTGACGCAACCTGCTGCTCCGTAACATCTGGGTACTGTCAACGAAAGACGTGAT 496  
DB      241 CGTGACGCAACCTGCTGCTCCGTAACATCTGGGTACTGTCAACGAAAGACGTGAT 300  
QY      497 AGCGTTGAAGCTACAACTTCATGTTGGAACTGGACATCAACAGCCGTTGATGATC 556  
DB      301 AGCGTTGAAGCTACAACTTCATGTTGGAACTGGACATCAACAGCCGTTGATGATC 360  
QY      557 GGCCTGATGGGAGGAGTCAAGCTGCGCGGACGCGAGTTCCGCCATGCTTCACTTC 616  
DB      361 GGCCTGATGGGAGGAGTCAAGCTGCGCGGAGGAGTTCCGCCATGCTTCACTTC 420  
QY      617 TGTGAAACCCGAGCAAAATTCAGTCCGTCGTGAACCTGCTGAAGAAATGACCGGCGT 676  
DB      421 TGTGAAACCCGAGCAAAATTCAGTCCGTCGTGAACCTGCTGAAGAAATGACCGGCGT 480  
QY      677 GAGCTGAACCGGACGACCGTACTGTAGACGTGACGATCGCCGTATTCGTAACATTTTC 736  
DB      481 GAGCTGAACCGGACGACCGTACTGTAGACGTGACGATCGCCGTATTCGTAACATTTTC 540  
QY      737 GAATCTAGCCCGGATAGCCCGGAA 760  
DB      541 GAATCTAGCCCGGATAGCCCGGAA 564
```

Search completed: January 22, 2005, 13:53:24  
Job time : 4273 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 22, 2005, 10:44:08 ; Search time 551 Seconds  
(without alignments)  
831.603 Million cell updates/sec

Title: US-10-613-990a-31

Perfect score: 927  
Sequence: 1 gcatcgtacgcgcgtacatcgtcgtt 927

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_23sep04:\*  
2: geneseqn1980s:\*  
3: geneseqn1990s:\*  
4: geneseqn2000s:\*  
5: geneseqn2001s:\*  
6: geneseqn2002s:\*  
7: geneseqn2003s:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	927	100.0	927	12	ADJ38402
2	925.8	99.9	1732	6	ABL90421
3	577.2	62.3	726	11	ACH96017
4	510.6	55.1	110000	10	ACF67367_28
5	510.6	55.1	110000	10	ACF65386_0
6	508.6	54.9	759	12	ADJ38380
7	493	53.2	717	10	ACF69844
8	450.8	48.6	726	10	ADP01748
9	447.4	48.3	1048	5	AAS29239
10	319.4	34.5	110000	2	AAT42063_09
11	316.2	34.1	711	4	AAS3409
12	316.2	34.1	711	8	ACA34287
13	232.8	25.1	1110	5	AAS82422
14	137.4	14.8	729	6	ABQ90117
15	126	13.6	837	11	ABD17813
16	126	13.6	2430	11	ABD17567
17	126	13.6	3675	11	ABD17705
18	120.6	13.0	990	11	ABD12913
19	120.6	13.0	2166	11	ABD13118
20	113.8	12.3	2004	11	ABD13261
21	107.4	11.6	1018	3	AAA38554

C	22	106.8	11.5	11275	6	AAH43717	Aah43717 E. coli g
	23	103.8	11.2	714	8	ACA47574	Aca47574 Prokaryot
	24	101.4	10.9	705	8	ACA48118	Aca48118 Prokaryot
	25	100	10.8	100	8	ACD69947	Acdd69947 E. coli K
	26	100	10.8	100	8	ACD69946	Acdd69946 E. coli K
	27	100	10.8	100	8	ACD69948	Acdd69948 E. coli K
	28	99	10.7	732	8	ACA26627	Aca26627 Prokaryot
	29	98.6	10.6	110000	6	ABA03041_03	Continuation (4 of
	30	98	10.6	714	8	ACA36700	Aca36700 Prokaryot
	31	97.2	10.5	762	11	ABD07703	Abd07703 Pseudomon
	32	96.8	10.4	349980	6	ABQ81844	Abq81844 Bifidobac
	33	96.2	10.4	705	6	ABN90763	Abn90763 Staphyloc
	34	96.2	10.4	721	8	ACA46767	ACA46767 Prokaryot
	35	95	10.2	110000	6	ABQ69245_03	Continuation (3 of
	36	95	10.2	110000	6	ABQ67197_02	Continuation (4 of
	37	94.6	10.2	702	4	AAH52543	Aah52543 S. epider
	38	94.6	10.2	3389	4	AAH54550	Aah54550 S. epider
	39	93	10.0	699	4	AAS52039	Aas52039 Staphyloc
	40	93	10.0	702	3	AAH55530	Aah55530 S. aureus
	41	93	10.0	708	4	AAS54939	Aas54939 Staphyloc
	42	93	10.0	744	8	ACA25690	Aca25690 Prokaryot
	43	92.6	10.0	792	11	ABD11939	Abd11939 Pseudomon
	44	92.6	10.0	804	11	ABD11719	Abd11719 Pseudomon
	45	91.4	9.9	699	4	AAS00812	Aas00812 S. aureus

#### ALIGNMENTS

RESULT 1	ADJ38402	standard; DNA; 927 BP.
ID	ADJ38402;	
XX	ADJ38402;	
AC	ADJ38402;	
XX	ADJ38402;	
DT	06-MAY-2004	(first entry)
XX	06-MAY-2004	(first entry)
DE	Escherichia coli arca (aerobic respiration control) gene SeqID31.	
XX	Escherichia coli arca (aerobic respiration control) gene SeqID31.	
KW	gamma-proteobacterium; Arca; aerobic respiration control;	
KW	Arca protein function; chromosomal arca gene; fermentation; L-amino acid;	
KW	L-lysine; L-glutamic acid; gene; de.	
XX	Escherichia coli.	
OS	Escherichia coli.	
XX	Escherichia coli.	
FH	Key	Location/Qualifiers
FT	CDS	101..817
FT	CDS	/*tag= a
XX	CDS	/product= "Escherichia coli arca protein"
PN	EP1382666-AI.	
XX	EP1382666-AI.	
PD	21-JAN-2004.	
XX	21-JAN-2004.	
PF	11-JUL-2003; 2003EP-00015911.	
XX	11-JUL-2003; 2003EP-00015911.	
PR	12-JUL-2002; 2002JP-00203764.	
XX	12-JUL-2002; 2002JP-00203764.	
PA	(AJIN ) AJINOMOTO CO INC.	
XX	(AJIN ) AJINOMOTO CO INC.	
PI	Ishikawa Y, Imaizumi A, Matsui K, Kojima H;	
XX	Ishikawa Y, Imaizumi A, Matsui K, Kojima H;	
DR	WPI; 2004-192975/19.	
XX	WPI; 2004-192975/19.	
DR	P-PSDB; ADJ38403.	
XX	P-PSDB; ADJ38403.	
PT	New bacterium useful for producing a target substance (especially L-	
XX	lysine, L-glutamic acid or another L-amino acid) is modified so that the	
PT	Arca protein does not normally function.	
XX	Arca protein does not normally function.	
PS	Claim 6; SEQ ID NO 31, 38pp; English.	
XX	Claim 6; SEQ ID NO 31, 38pp; English.	
CC	This invention relates to a novel gamma-proteobacterium able to produce a	
XX	target substance such as an amino acid which is modified so that an Arca	

CC (aerobic respiration control) protein does not normally function. The  
CC disruption to Arca protein function is preferably due to disruption of a  
CC chromosomal arca gene. The bacterium and method are useful in producing a  
CC target substance by fermentation. The target substance is preferably an L  
CC -amino acid, especially L-lysine or L-glutamic acid. Production of the  
CC target substance using the modified bacterium is more efficient. The  
CC present sequence is that of a PCR primer which was used for amplification  
CC of the P ananatis arca gene in the exemplification of the invention.  
XX

Sequence 927 BP; 236 A; 228 C; 225 G; 238 T; 0 U; 0 Other;

Query Match 100.0%; Score 927; DB 12; Length 927;  
Best Local Similarity 100.0%; Pred. No. 1.1e-257;  
Matches 927; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTCATGTTAGCCGATCATGTTAAATTGACATGATCAGGAGGAGTCAAGGACTTTTGT 60
DB 1 GTCATGTTAGCCGATCATGTTAAATTGACATGATCAGGAGGAGTCAAGGACTTTTGT 60
QY 61 ACTTCCTGTTGATTTAGTTAGTGGCAATTTAGTACAAACATGCAACCCCGACATTTCT 120
DB 61 ACTTCCTGTTGATTTAGTTAGTGGCAATTTAGTACAAACATGCAACCCCGACATTTCT 120
QY 121 TATCGTTGAAGACAGTTGTTGAACACGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
DB 121 TATCGTTGAAGACAGTTGTTGAACACGCAACAGTTGAAAGTATTTTGAAGCGGAAG 180
QY 181 CTATGATGTTTTCAGACGACAGATGCGCGGAAATGATCAGATCCTCTCTGAATATGA 240
DB 181 CTATGATGTTTTCAGACGACAGATGCGCGGAAATGATCAGATCCTCTCTGAATATGA 240
QY 241 CATCAACCTGATGATCATGATGATCATGCGCGGATGAAAGAGTCTTCTTTAGCGCG 300
DB 241 CATCAACCTGATGATCATGATGATCATGCGCGGATGAAAGAGTCTTCTTTAGCGCG 300
QY 301 TGAACCTGCGGAGACGAGGAGTGTGCTGATGTTCTGATCGCGCGTGAACAAGAT 360
DB 301 TGAACCTGCGGAGACGAGGAGTGTGCTGATGTTCTGATCGCGCGTGAACAAGAT 360
QY 361 CGATTAATAATCTGCGGCTCGAATCGGTGAGATGATCATCAACCAACCGTTCAACC 420
DB 361 CGATTAATAATCTGCGGCTCGAATCGGTGAGATGATCATCAACCAACCGTTCAACC 420
QY 421 GCGTGAACCTGACGATTCGTCAGCAACCTAAGTCCCGATCAAGTAACTGGTATCT 480
DB 421 GCGTGAACCTGACGATTCGTCAGCAACCTAAGTCCCGATCAAGTAACTGGTATCT 480
QY 481 CAGCGAAGAACCTGTCAGCTGTAAGCTTAAGTCAATGATGTTGGAACCTGGAATCA 540
DB 481 CAGCGAAGAACCTGTCAGCTGTAAGCTTAAGTCAATGATGTTGGAACCTGGAATCA 540
QY 541 CAGCGCTTGTGATCGGCTCGATGAGGAGAGTAAAGCTGCGCGGAGGAGTTCG 600
DB 541 CAGCGCTTGTGATCGGCTCGATGAGGAGAGTAAAGCTGCGCGGAGGAGTTCG 600
QY 601 CCGCATGCTTCACTTCTGTAAGAACCCAGGCAAAATGATGCTGCTGAACCTGTA 660
DB 601 CCGCATGCTTCACTTCTGTAAGAACCCAGGCAAAATGATGCTGCTGAACCTGTA 660
QY 661 GAAATGACCGGCGGAGCTGAAGCCGACGACCTGTAAGCTGTAAGCTGATCGCG 720
DB 661 GAAATGACCGGCGGAGCTGAAGCCGACGACCTGTAAGCTGTAAGCTGATCGCG 720
QY 721 TATTCGTAACATTTGATCTACCGCGGATACGCGGAAATCAATCGCACATTTCA 780
DB 721 TATTCGTAACATTTGATCTACCGCGGATACGCGGAAATCAATCGCACATTTCA 780
QY 781 TGAAGTATCGCTTCTGCGGATCTGAAAGATTAATCGCTTACCAACCGTCAAAAA 840
DB 781 TGAAGTATCGCTTCTGCGGATCTGAAAGATTAATCGCTTACCAACCGTCAAAAA 840
QY 841 AACGCGCTTTTAGCGCGCTTTTATTTTCAACCTATTTTCAAGTACGTAATCATC 900
DB 841 AACGCGCTTTTAGCGCGCTTTTATTTTCAACCTATTTTCAAGTACGTAATCATC 900
```

DB 841 AACGCGCTTTTAGCGCGCTTTTATTTTCAACCTATTTTCAAGTACGTAATCATC 900  
QY 901 GTCCGTTGTAACCTTTTACTGCTTT 927  
DB 901 GTCCGTTGTAACCTTTTACTGCTTT 927

## RESULT 2

ABL90421/c  
ID ABL90421 standard; cDNA; 1732 BP.

AC ABL90421;  
XX

DT 24-MAY-2002 (first entry)  
XX

DE Human polynucleotide SEQ ID NO 983.  
XX

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein; gene; ss.

OS Homo sapiens.  
XX

PN WO200190304-A2.  
XX

XX 29-NOV-2001.  
XX

PF 18-MAY-2001; 2001WO-US016450.  
XX

PR 19-MAY-2000; 2000US-0205515P.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.  
XX

PI Birse CE, Rosen CA;  
XX

XX WPI; 2002-122018/16.  
XX

DR P-PSDB; ABB90012.  
XX

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders.

XX Claim 4; SEQ ID NO 983; 2081bp + Sequence Listing; English.

XX The invention relates to novel genes (ABL9449-ABL90853) and proteins  
XX (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
XX medical conditions e.g. by protein or gene therapy. The genes are  
XX isolated from a range of human tissues disclosed in the specification.  
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
XX ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
XX breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
XX disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
XX anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
XX multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
XX cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
XX / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
XX infections. Note: The sequence data for this patent did not form part of  
XX the printed specification, but was obtained in electronic format directly  
XX from WIPo at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 1732 BP; 439 A; 391 C; 417 G; 477 T; 0 U; 8 Other;

Query Match 99.9%; Score 925.8; DB 6; Length 1732;  
Best Local Similarity 99.7%; Pred. No. 3.9e-297;  
Matches 924; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTCATGTTAGCCGATCATGTTAAATTGACATGATCAGGAGGAGTCAAGGACTTTTGT 60
|||
```



Db 1313 GTCATGTACGCCGATCATTTAATTGACATGCATCAGGAGTCCGAGACTTTTGT 1254  
QY ACTTCCTGTTTCGATTTAGTGGCAATTAGTAGCAAAATGATGAGACCCCGCATCTT 120  
Db 1253 ACTTCCTGTTTCGATTTAGTGGCAATTAGTAGCAAAATGATGAGACCCCGCATCTT 1194  
QY 121 TATGTTGAAGACGAGTGGTAAACAGCAACAGCTGAAAGTATTTTGAAGCGAAG 180  
Db 1193 TATGTTGAAGACGAGTGGTAAACAGCAACAGCTGAAAGTATTTTGAAGCGAAG 1134  
QY 181 CTATGATGTTTGAAGCGACAGATGGCGGAAATGCAATCATCTCTCTGAATTTGA 240  
Db 1133 CTATGATGTTTGAAGCGACAGATGGCGGAAATGCAATCATCTCTCTGAATTTGA 1074  
QY 241 CATCAACCTGATGATCATGATATCAATCTCCGGGTAAAGACGCTTCTGTTAGCGC 300  
Db 1073 CATCAACCTGATGATCATGATATCAATCTCCGGGTAAAGACGCTTCTGTTAGCGC 1014  
QY 301 TGAATGCGCGAGACGAGCGAATGTTGCTTATGTTCTGATCTGACCTGACCAACGAA 360  
Db 1013 TGAATGCGCGAGACGAGCGAATGTTGCTTATGTTCTGATCTGACCTGACCAACGAA 954  
QY 361 CGATTAATTTCTGGCTTCCGTAATTCGCTGAGATGATCATCATCAACCAACCGTTCAACC 420  
Db 953 CGATTAATTTCTGGCTTCCGTAATTCGCTGAGATGATCATCATCAACCAACCGTTCAACC 894  
QY 421 GCGTGAATGACGATTCGTGACAGCAACCTACTGTCCTGACCATGATGATGATGATGAT 480  
Db 893 GCGTGAATGACGATTCGTGACAGCAACCTACTGTCCTGACCATGATGATGATGATGAT 834  
QY 481 CAGCGAAGAACGTCGTAGCGTTGAAAGCTACAGATGATGATGATGATGATGATGATGAT 540  
Db 833 CAGCGAAGAACGTCGTAGCGTTGAAAGCTACAGATGATGATGATGATGATGATGATGAT 774  
QY 541 CAGCGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600  
Db 773 CAGCGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 714  
QY 601 CCGCATGCTTCACTTCTGTGAAACCAAGCAAAATTCAGTCCCGTCTGATGATGATGAT 660  
Db 713 CCGCATGCTTCACTTCTGTGAAACCAAGCAAAATTCAGTCCCGTCTGATGATGATGAT 654  
QY 661 GAAATGACCGCGCGTGAAGCGCAACCGTACTGATGATGATGATGATGATGATGATGAT 720  
Db 653 GAAATGACCGCGCGTGAAGCGCAACCGTACTGATGATGATGATGATGATGATGATGAT 594  
QY 721 TATTCGTAACATTTGGAATCTACCGCGGATACCGCGGAAATCATGCGCCATTCACGG 780  
Db 593 TATTCGTAACATTTGGAATCTACCGCGGATACCGCGGAAATCATGCGCCATTCACGG 534  
QY 781 TGAAGGTTATCGCTTCTGCGGATCTGAAAGATTAATGCGCTTTACCAACCGTCAAAAA 840  
Db 533 TGAAGGTTATCGCTTCTGCGGATCTGAAAGATTAATGCGCTTTACCAACCGTCAAAAA 474  
QY 841 AACGGGCTTTTGAAGCGCGCTTTTATTTTCAACCTTATTTTCAGATGATGATGATGAT 900  
Db 473 AACGGGCTTTTGAAGCGCGCTTTTATTTTCAACCTTATTTTCAGATGATGATGATGAT 414  
QY 901 GTCCGTTTGAATCTTTTACTGCGCTT 927  
Db 413 GTCCGTTTGAATCTTTTACTGCGCTT 387

RESULT 3  
ACH96017  
ID ACH96017 standard; DNA: 726 BP.  
XX ACH96017;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polynucleotide seqid 1812.  
XX

KW Recombinant expression vector; transcription regulatory element;  
KM Klebsiella pneumoniae protein; antibacterial; Vaccine; gene; ds.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN US6610836-B1.  
XX  
PD 26-AUG-2003.  
XX  
PF 27-JAN-2000; 2000US-00489039.  
XX  
PR 29-JAN-1999; 99US-0117747P.  
XX  
XX (GENO-) GENOME THERAPEUTICS CORP.  
PI Breton GL, Osborne M;  
PI WPI; 2003-895346/82.  
DR P-PSDB; AB062466.  
XX  
XX New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.  
XX  
XX Disclosure; SEQ ID NO 1812; 932pp; English.  
XX  
XX The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid; operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This sequence encodes a Klebsiella  
CC pneumoniae polypeptide of the invention  
XX  
SQ Sequence 726 BP; 189 A; 198 C; 189 G; 150 T; 0 U; 0 Other;  
Query Match 62.3%; Score 577.2; DB 11; Length 726;  
Best Local Similarity 87.2%; Pred. No. 3,7e-181;  
Matches 633; Conservative 0; Mismatches 93; Indels 0; Gaps 0;  
QY 92 GTAGCAACATGAGAGCCCGGACATTTTATGTTGAAGAGAGTGGTAACAGCAAC 151  
Db 1 GTAGCAACATGAGAGCCCGGACATTTTATGTTGAAGAGAGTGGTAACAGCAAC 60  
QY 152 ACCTGAAAGATTTTTCGAAGCGAAGCTATGATGTTTTCGAAGCGACAGATGCGCG 211  
Db 61 ACCTGAAAGATTTTTCGAAGCGAAGCTATGATGTTTTCGAAGCGACAGATGCGCG 120  
QY 212 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 271  
Db 121 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 272 CCGGTTAAGAACGCTTCTGTTAGCGCGTGAATCTCCGAGAGAGCGAATGTTGGTTG 331  
Db 181 CCGGTTAAGAACGCTTCTGTTAGCGCGTGAATCTCCGAGAGAGCGAATGTTGGTTG 240  
QY 332 ATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 391  
Db 241 ATGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 392 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 451  
Db 301 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 452 CTGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511  
Db 361 CTGTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
QY 512 AAGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 571  
Db 421 AAGTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
QY 572 CAGTACAGCTGCGGAGAGAGTTCGCGGATGATGATGATGATGATGATGATGATGATGAT 631

Db 481 CAGTACAGCTCCGCGCAGTGAATTCGCGCGATGCTGACCTTTCGCAAAACCCGGCC 540  
Qy 632 AAATTCAGTCCCGTGTGAAGCTGAATAAATGACCGCGCTGAGCTGAAACCGCAC 691  
Db 541 AAATTCAGTCCCGTGTGAAGCTGAATAAATGACCGCGCTGAGCTGAAACCGCAC 600  
Qy 692 GACCGTACTGTAAGCTGAGATCCGCGTATTCGTAAACATTTGCAATCTACGCCGAT 751  
Db 601 GACCGTACCGTTACGTAACATTCGCGGATTCGTAAACATTCGCAATCCACCTCCGAT 660  
Qy 752 ACCCGGAATCATCCGACCATTCACGCTGAAGGTTATCGCTTCCGCTGATCTGAA 811  
Db 661 ACCCGGAATCATTCGTAACATTCACGCTGAAGGTTATCGCTTCCGCTGATCTGAA 720  
Qy 812 GATTA 817  
Db 721 GATTA 726

RESULT 4  
ACF67367\_28  
Continuation (29 of 57) of ACF67367 from base 2800001 (Photophobus luminescens nucleoti  
WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367  
Fragment Name Begin End  
WP ACF67367\_00 1 110000  
WP ACF67367\_01 100001 210000  
WP ACF67367\_02 200001 310000  
WP ACF67367\_03 300001 410000  
WP ACF67367\_04 400001 510000  
WP ACF67367\_05 500001 610000  
WP ACF67367\_06 600001 710000  
WP ACF67367\_07 700001 810000  
WP ACF67367\_08 800001 910000  
WP ACF67367\_09 900001 1010000  
WP ACF67367\_10 1100000 1110000  
WP ACF67367\_11 1100001 1210000  
WP ACF67367\_12 1200001 1310000  
WP ACF67367\_13 1300001 1410000  
WP ACF67367\_14 1400001 1510000  
WP ACF67367\_15 1500001 1610000  
WP ACF67367\_16 1600001 1710000  
WP ACF67367\_17 1700001 1810000  
WP ACF67367\_18 1800001 1910000  
WP ACF67367\_19 1900001 2010000  
WP ACF67367\_20 2000001 2110000  
WP ACF67367\_21 2100001 2210000  
WP ACF67367\_22 2200001 2310000  
WP ACF67367\_23 2300001 2410000  
WP ACF67367\_24 2400001 2510000  
WP ACF67367\_25 2500001 2610000  
WP ACF67367\_26 2600001 2710000  
WP ACF67367\_27 2700001 2810000  
WP ACF67367\_28 2800001 2910000  
WP ACF67367\_29 2900001 3010000  
WP ACF67367\_30 3000001 3110000  
WP ACF67367\_31 3100001 3210000  
WP ACF67367\_32 3200001 3310000  
WP ACF67367\_33 3300001 3410000  
WP ACF67367\_34 3400001 3510000  
WP ACF67367\_35 3500001 3610000  
WP ACF67367\_36 3600001 3710000  
WP ACF67367\_37 3700001 3810000  
WP ACF67367\_38 3800001 3910000  
WP ACF67367\_39 3900001 4010000  
WP ACF67367\_40 4000001 4110000  
WP ACF67367\_41 4100001 4210000  
WP ACF67367\_42 4200001 4310000  
WP ACF67367\_43 4300001 4410000  
WP ACF67367\_44 4400001 4510000  
WP ACF67367\_45 4500001 4610000  
WP ACF67367\_46 4600001 4710000  
WP ACF67367\_47 4700001 4810000  
WP ACF67367\_48 4800001 4910000

WP ACF67367\_49 4900001 5010000  
WP ACF67367\_50 5000001 5110000  
WP ACF67367\_51 5100001 5210000  
WP ACF67367\_52 5200001 5310000  
WP ACF67367\_53 5300001 5410000  
WP ACF67367\_54 5400001 5510000  
WP ACF67367\_55 5500001 5610000  
WP ACF67367\_56 5600001 5648894

Query Match 55.1%; Score 510.6; DB 10; Length 110000;  
Best Local Similarity 77.5%; Pred. No. 1e-157;  
Matches 618; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

Qy 44 AGGTACAGGACTTTTGTACTCTCTTTCGATTTAGTTAGTGGCAATTTAGTGAACATG 103  
Db 67943 AGTAAAGAGTATTTGAATCATTCGATTTATTTTAAAGCATTTCTAGTGAACATG 68002  
Qy 104 CAGACCCCGACATTTCTTATCGTTGAGACGAGTGTGAACGCAACACGTTGAAAGT 163  
Db 68003 CAAACCCCGACATTTGATTTGTTGAAGCAAAATGTCACCTCCCAATCCCTTAAAGC 68062  
Qy 164 ATTTTGAAGCGAAGCTATGATTTTTCGACGACGATGCGCGAAATGATCAG 223  
Db 68063 ATTTTGAAGCTGAAGGCTATGATTTATGAAGCCATGATGTTCAAGAAATGACCAT 68122  
Qy 224 ATCTCTCTGAATATGATCATCACTGTGATCATGATATCATCTGCGGGTAAAGAC 283  
Db 68123 ATCTGTCAAAACAAGCATTAACCTGTGTATGATATTAACCTTCCAGGCAAAAT 68182  
Qy 284 GATCTTCTGTTAGGCGGTGAACGCGCGACAGCCGATTTTCGTTGATGTTCTGACT 343  
Db 68183 GCTTTGTTACTTCTGCGCGAATGCGAGCAAGCAAAATGTTCTTGTGATGTTCTGACT 68242  
Qy 344 GCGCTGACAAAGAGTGAATAAATCTCGGCTCGAATCGGTGCAATGATCATC 403  
Db 68243 GCGCTGATTAACGAAGTGAACAAATCTGCTTGGAAATCGGTGCAATGATCATC 68302  
Qy 404 ACCAAACGTTCAACCGCGGTGAACGAGATTCGTGACGCAACCTACTGTCCTGACC 463  
Db 68303 ACCAAACATTTTAATCGCGTGAATGACTATTCGCGCGCAACCTGCTTCTGTAAC 68362  
Qy 464 ATGAATCTGGATCTGTCAAGCAAGACGTGAGCGTTGAAGCTACAAAGTTCAATGAT 523  
Db 68363 ATGAATCTGATGATGTCAAGCAAGACCTGCCAAGTTGAGATTTAATTAATCAACGT 68422  
Qy 524 TGGAACTGACATCAACAGCGCTTGTGATGCGCCCTGATGCGACAGTAAACAGCTG 583  
Db 68423 TGGAAATTAATATCAATGACCGTTCCTGATTAAGCCCGGCTGTGACCGATTAATG 68482  
Qy 584 CCGCGAGGAGATTCCGCGCATGCTTCACTCTGTAACCCAGGCAAAATCAATGATCC 643  
Db 68483 CCAAGGAGGAGATTCCGCGCATGCTTCACTCTGTAACCCAGGCAAAATCAACAG 68542  
Qy 644 CGTCTGAATCTGTGAAGAAATGACCGCGGTGAGTGAACGCAACGCTACTGTA 703  
Db 68543 CGCGGATTTAATCTGAAGAAATGACAGGTGTAATGAACCTCAAGACCGTACTGTT 68602  
Qy 704 GACGTGAGATCCGCGTATTCGTAACATTTGAAATCTACCGGATACCGCGAAATC 763  
Db 68603 GATGTAAGATTCGTCGATTCGTAACATTTGAAATCTACCGGATACCGCGAAATC 68662  
Qy 764 ATCGGACCATTTACGCGTGAAGGTTATCGCTTTCGCGTATCTGAAGATTAATGCGCT 823  
Db 68663 ATCGGACCATTTACGCGTGAAGGCTATCGTTCTGTGTATTTGAAAGATTAACCTGAA 68722  
Qy 824 TTACACCGTCAAAAA 840  
Db 68723 CTATGCTCTTAACA 68739

RESULT 5  
ACF65386\_0/c  
WP Sequence split into 7 fragments LOCUS ACF65386 Accession ACF65386

WP	Fragment Name	Begin	End
WP	ACF65386_0	1	110000
WP	ACF65386_1	100001	210000
WP	ACF65386_2	200001	310000
WP	ACF65386_3	300001	410000
WP	ACF65386_4	400001	510000
WP	ACF65386_5	500001	610000
WP	ACF65386_6	600001	700779
ID	ACF65386 standard; DNA; 700779 BP.		
AC	ACF65386;		
XX	20-NOV-2003 (first entry)		
DE	Photorhabdus luminescens nucleotide sequence #39.		
XX	Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;		
KM	detection; food; gene expression; plant; animal; microorganism; toxin;		
KM	antibiotic; biopesticide; virulence factor; disease model; plague;		
XX	whooping cough; gene; ds.		
OS	Photorhabdus luminescens.		
PN	MO200294867-A2.		
PD	28-NOV-2002.		
PF	07-FEB-2002; 2002MO-IB003040.		
PR	07-FEB-2001; 2001FR-00001659.		
XX	(INSP ) INST PASTEUR.		
PA	(CNRS ) CNRS CENT NAT RECH SCI.		
PI	Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunet F, Danchin A;		
PI	Buchreiser C;		
DR	WPI; 2003-148459/14.		
XX	Genomic sequence of Photorhabdus luminescens and encoded polypeptides;		
PT	useful e.g. as therapeutic antimicrobials and agricultural pesticides.		
XX	Claim 1; SEQ ID NO 39; 1205bp; French.		
PS	The invention relates to the isolation of genes and their encoded		
CC	proteins from Photorhabdus luminescens. The isolated sequences are		
CC	sources of probes and primers for detecting the genome of P. luminescens		
CC	and related species; to study polymorphisms; for gene analysis and for		
CC	detection/amplification of the genes. Antibodies (Ab) raised against the		
CC	polypeptides encoded by the genes are used for detection/identification		
CC	of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that		
CC	carry a gene-containing vector are used to select compounds that		
CC	modulate, regulate, induce or inhibit expression of the genes in plants,		
CC	animals or microorganisms other than P. luminescens and are able to alter		
CC	response or sensitivity to toxins and antibiotics produced by P.		
CC	luminescens. Cells transformed to express the genes are useful for		
CC	recombinant production of the proteins, particularly toxins and		
CC	antibacterials useful as insecticides, bactericides and fungicides. The		
CC	genes, proteins, vectors containing the genes and Ab are also useful		
CC	therapeutically (to treat microbial infection by bacteria or fungi that		
CC	are sensitive to P. luminescens-encoded toxins or antibiotics) and as		
CC	biopesticides. Other uses of the genes and the proteins are as virulence		
CC	factors and for identifying targets of human diseases for which P.		
CC	luminescens is a model (particularly plague and whooping cough). This		
CC	sequence represents one of the isolated P. luminescens genes		
XX			
SQ	Sequence 700779 BP; 190440 A; 140977 C; 164444 G; 204917 T; 0 U; 1 Other;		
Query Match	55.1%; Score 510.6; DB 10; Length 110000;		
Best Local Similarity	77.5%; Pred. No. 1e-157.		
Matches	618; Conservative 0; Mismatches 179; Indels 0; Gaps 0;		
QY	44 AGGTACGAGCACTTTGTTACTCTGTTTCGATTAGTTCGCAATTAGTAGCAAAACATG 103		

Db	30684	AGATAAGAGTAGTTTAATCAATTCGATTATTTTAACGATTTAGTAGCAACCATG 30625
QY	104	CAGACCCCGCACATTTCTTATCGTTGAAGAGAGTTGTAAACGCAACAGTTGAAAGT 163
Db	30624	CAGACCCCGCACATTTCTTATCGTTGAAGAGAGTTGTAAACGCAACAGTTGAAAGT 30565
QY	164	ATTTTGAAGCGGAGCTATGATGTTTTCGAAAGCAAGATGCGCGGAAATGATCAG 223
Db	30564	ATTTTGAAGCTGAAGAGGTACATAGTTATGAAGCCATGATGTTTTCAGAAATGACCAT 30505
QY	224	ATCCTCTGTAATATATCAATCAACCTGTGATATCATATGATATGATGATGATGATG 283
Db	30504	ATTCGTCAAAACACACATTAACCTGTGATATGATATGATATGATGATGATGATGATG 30445
QY	284	GGCTCTCTGTTAGCGGTGAACCTGCGAGACGAGCGCAATGTTGCTGTGATGTTCTGACT 343
Db	30444	GGCTTGTACTTCCCGGCACTGCGAGAGCAAGCAAAATGTTGCTGTGATGTTCTGACT 30385
QY	344	GGCGGTGACAAAGCAAGTGAATAAATTCGCGCTCGAAATCGGTGCAATGATGATCATC 403
Db	30384	GGCGGTGATTAAGCAAGTGAATAAATTCGCGCTCGAAATCGGTGCAATGATGATCATC 30325
QY	404	ACCAACCGTTCAACCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 463
Db	30324	ACCAACCAATTTAATCCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATG 30265
QY	464	ATGATCTGGGTACTGTCAGCAAGCAAGCGTGAAGGTGAAGTGAAGTGAAGTGAAGTGA 523
Db	30264	ATGATCTGAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 30205
QY	524	TGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 583
Db	30204	TGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 30145
QY	584	CCGCGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 643
Db	30144	CCAGCAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 30085
QY	644	CGTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 703
Db	30084	CGGCGATTTTACTGTAATAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 30025
QY	704	GACGTGACATCGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 763
Db	30024	GATGTAAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 29965
QY	764	ATGCGACATTCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 823
Db	29964	ATGCGACATTCAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 29905
QY	824	TTACGACCGTCAAAAAA 840
Db	29904	CTATCGGTCTTAACA 29888
RESULT 6		
ID	ADJ38390	standard; DNA; 759 BP.
AC	ADJ38390;	
XX	06-MAY-2004 (first entry)	
DE	Pantoea ananatis arca (aerobic respiration control) gene SeqID19.	
XX		
KM	gamma-proteobacterium; Arca; aerobic respiration control;	
KM	Arca protein function; chromosomal arca gene; fermentation; L-amino acid;	
KM	L-lysine; L-glutamic acid; gene; ds.	
OS	Pantoea ananatis.	
XX		
FH	Key	Location/Qualifiers

FT CDS 41..757  
 FT /\*tag= a  
 FT /product= "Pantoea ananatis arca protein"  
 PN EPI382686-A1.  
 PD 21-JAN-2004.  
 XX 11-JUL-2003; 2003EP-00015911.  
 XX 12-JUL-2002; 2002JP-00203764.  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA Iehikawa Y, Imaizumi A, Matsui K, Kojima H;  
 PI WPI; 2004-192975/19.  
 DR P-PSDB; ADJ38391.  
 XX  
 XX New bacterium useful for producing a target substance (especially L-  
 PT lysine, L-glutamic acid or another L-amino acid) is modified so that the  
 PT Arca protein does not normally function.  
 XX  
 PS Example 4; SEQ ID NO 19; 38pp; English.  
 XX  
 CC This invention relates to a novel gamma-proteobacterium able to produce a  
 CC target substance such as an amino acid which is modified so that an Arca  
 CC (aerobic respiration control) protein does not normally function. The  
 CC disruption to Arca protein function is preferably due to disruption of a  
 CC chromosomal arca gene. The bacterium and method are useful in production of a  
 CC target substance by fermentation. The target substance is preferably an L-  
 CC amino acid, especially L-lysine or L-glutamic acid. Production of the  
 CC target substance using the modified bacterium is more efficient. The  
 CC present sequence is that of the P ananatis gene which encodes the Arca  
 CC protein and which was used in the exemplification of the invention.  
 CC  
 SQ Sequence 759 BP; 205 A; 192 C; 183 G; 179 T; 0 U; 0 Other;

Query Match 54.9%; Score 508.6; DB 12; Length 759;  
 Best Local Similarity 80.6%; Pred. No. 2.8e-158;  
 Matches 607; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

QY 69 TTTGATTGTTGGTGGC--AATTAGTAGCAAAATGATGACACCCGCGCATTTTATGCT 126  
 Db 7 TTTCAATTATGTTGGCAAAATTAGTAGTAAACATGACAGACCCGCGCATTTCTATGCT 66  
 QY 127 TGAAGACGAGTTGGTGAACGCAACGTTGAAAGTATTTTGAAGCCGAAGGCTATGA 186  
 Db 67 TGAAGACGAGTGGTGAACGCGCAATACCTCAAAAGCATTTTGAAGGCGGAAGTTATGT 126  
 QY 187 TGTTCGGAAGGACAGATGGCGGGAATGATCAGATCTCTGGAATATGATCAATCAA 246  
 Db 127 CGGTACGAAAGCAGATGGTGAGATGACAGCAGGTTGACCGACGATGATGTCAA 186  
 QY 247 CCTGTGATCATGATATCAATCTGCGGGTAAAGACGCTCTCTGTTAGCGCGTAACT 306  
 Db 187 TCTGTTATATGAGATCAATCTGCGGGTAAAGACGCTCTCTGTTAGCGCGTAACT 306  
 QY 307 GGGCGGACGCGGAATGTTGCGTTGATGTTCTGACTGCGCGTGAACAAGTCGATGA 366  
 Db 247 GCGTGAAGGCAATGTCGATGATGTTCTGACCGGACGATTAAGAAAGTCGATGA 306  
 QY 367 AATTCGCGCTCGAAATGCGTGAAGATGATCACTACCAACCGTTCAACCGCGTGA 426  
 Db 307 AATTCGCGCTCGAAATGCGTGAAGATGATCACTACCAACCGTTCAACCGCGTGA 426  
 QY 427 ACTGACGATTCGTGACGCAACCTACTGCTCCGTAACATGAAATGCGGTACTGTCAGCA 486  
 Db 367 ATTAACTATTGTCGACGCTAAGCTGCTTGGCGACATGATTAATTTGCTTTACCCATGA 426  
 QY 487 AGAAGCTGATGCGTTGAAAGCTTACAGTCAATGTTGGGAATGAGATCAACAGCCG 546  
 Db 427 AGAGCTGCGCAAGTTGAAAGCTTAAAGTTCAACGCGCTGGAGCTGAGCATCAACAGCCG 486

QY 547 TTGCTTATGCGCCCTGATGCGAGCAAGTACAGCTGCCGCGCAGAGATTCCGCCCAT 606  
 Db 487 CTACATCATTAATCCCAAGCGTGAAGTAAACAGTCCGCGCAGTGAATTCGATGCGCAT 546  
 QY 607 GCTTCACTTCTGTGAAACCCAGGCAAAATTAGTCCGCTGTGAACGCTGAAGAAAT 666  
 Db 547 GCTGCACTTCTGTGAAACCCAGGCAAAATTAGTCCGCTGTGAACGCTGAAGAAAT 606  
 QY 667 GACCGGCGGTGAGCTGAAACCCAGCAGCTGATCTGTGAACGATCCGCGTATTCG 726  
 Db 607 GACCGGCGGTGAGCTGAAACCCAGCAGCTGATCTGTGAACGATCCGCGTATTCG 666  
 QY 727 TAAACATTTGATCTACGCGCGGATAGCGCGGAAATCTGCGCACATTCGCGTAAAG 786  
 Db 667 TAAACATTTGATCTACGCGCGGATAGCGCGGAAATCTGCGCACATTCGCGTAAAG 726  
 QY 787 TTAATGCTTTCTGCGGTGATCTGGAAGATTAAATC 819  
 Db 727 TTAATGCTTTCTGCGGTGATCTGGAAGATTAAATC 759

RESULT 7  
 ACF69844  
 ID ACF69844 standard; DNA; 717 BP.  
 AC ACF69844;  
 XX

DT 20-NOV-2003 (first entry)

DE Photorhabdus luminescens nucleotide sequence #8311.

KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
 KW detection; food; gene expression; plant; animal; microorganism; toxin;  
 KW antibiotic; biopesticide; virulence factor; disease model; plague;  
 KW whooping cough; gene; ds.

KM Photorhabdus luminescens.

OS WO200294867-A2.

PD 28-NOV-2002.

PF 07-FEB-2002; 2002MO-IB003040.

PR 07-FEB-2001; 2001FR-00001659.

PA (INSP ) INST PASTEUR.

PA (CNRS ) CNRS CENT NAT RECH SCI.

PI Duchaud E, Taourit S, Glaeser P, Frangeul L, Kunst F, Danchin A;  
 PI Buchrieser C;

DR WPI; 2003-148459/14.

PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
 XX  
 PS Claim 2; SEQ ID NO 8311; 1205bp; French.

CC The invention relates to the isolation of genes and their encoded  
 CC proteins from Photorhabdus luminescens. The isolated sequences are  
 CC sources of probes and primers for detecting the genome of P. luminescens  
 CC and related species; to study polymorphisms; for gene analysis and for  
 CC detection/amplification of the genes. Antibodies (Ab) raised against the  
 CC polypeptides encoded by the genes are used for detection/identification  
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
 CC carry a gene-containing vector are used to select compounds that  
 CC modulate, regulate, induce or inhibit expression of the genes in plants,  
 CC animals or microorganisms other than P. luminescens and are able to alter  
 CC response or sensitivity to toxins and antibiotics produced by P.  
 CC luminescens. Cells transformed to express the genes are useful for  
 CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as  
CC biopesticides. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which *P.*  
CC *luminescens* is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated *P. luminescens* genes

XX Sequence 717 BP; 213 A; 158 C; 158 G; 188 T; 0 U; 0 Other;

Query Match 53.2%; Score 493; DB 10; Length 717;  
Best Local Similarity 80.5%; Pred. No. 4.2e-153;  
Matches 577; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 101 ATGACAGACCCCGACATTTCTTATCGTTGAAGACGAGTTGGTAAACAGCAACAGCTTGAA 160  
DB 1 ATCCAAACCCCGACATTTTGAATTTGAAGACGAAATTTCTCTCCCAATACCTTA 60  
QY 161 AGTATTTTGAAGCGGAGGCTATGATGTTTGAACGACAGATGGCCGGAATGCAT 220  
DB 61 AGCATTTTGAAGCTGAAGGGTACATAGTTTATGAAGCCACTGATGGTTCAAGAAATGCAC 120  
QY 221 CAGATCTCTCTGAATATGATCAATCAACCTGTGATCATGATATCAATCTGCCGGTAG 280  
DB 121 CATATTTCTGCAAAACAAGCATTTAATCTGTGATATGATATTAACCTTCCAGGCA 180  
QY 281 AACGCTCTTCTGTAGCGCGTGAATGCGCGAGACGAGCGCAATGTTGGTTGATGTTCTG 340  
DB 181 MATGCTTTTACTTCTCCCGCAATCTCGAGACGACCAATTTTCTTTGATGTTCTG 240  
QY 341 ACTGCGCGTGAACAAGTGCATTAATTTCTGCGCTCGAAATCGGTGCAGATGATAC 400  
DB 241 ACTGCGCGTGAATACAGAGTGAACAATCTTGGTGGTGAATCGGTGCAGATGATAC 300  
QY 401 ATCACCAAAACGTTCAACCGCGGTGAATGACATTCGTGACAGCACTGATCTCCGT 460  
DB 301 ATCACCAAAACATTTAATCGCGTGAATGATGATTCGCGCCGCAACCTGTTCTCGT 360  
QY 461 ACCATGATCTGGGTACTGTCAGCAAGAAAGTGTGAGGTTGAAGCTCAAGTTCAAT 520  
DB 361 ACCATGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
QY 521 GGTGGGAATGACATCAACAGCCGTTGATGATGATGATGATGATGATGATGATGATGAT 580  
DB 421 GGTGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
QY 581 CTGCGCGGACGAGTTCGCGCATGCTTCACTTCTGTAACCAAGGCAAAATTCAG 640  
DB 481 TTGCGCAGCAGAGTTCGCGCATGCTTCACTTCTGTAACCAAGGCAAAATTCAG 540  
QY 641 TCCCGGTGGAATGCTGTAAGAAATGACCGCGCGTGAAGCTGAACCGCAACCGTACT 700  
DB 541 ACACGGCAGATTTACTGAAAGAAAGATGACAGTGTGAATGAACCTCAACCGTACT 600  
QY 701 GTAGACGTGACGATCCGCGTATTCGTAAACATTTGAAATCTAGCGCGGATAGCGGAA 760  
DB 601 GTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
QY 761 ATCATGCGCAACATTCAGGTTGAGGTTATCGCTTCTGCGGTGATCTGAAGATTA 817  
DB 661 ATTATGCGCACATTCAGGTTGAGGTTATCGCTTCTGCGGTGATCTGAAGATTA 717

RESULT 8  
ADP01748  
ID ADP01748 standard; DNA; 726 BP.  
XX  
AC ADP01748;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Bacterial polynucleotide #2033.

XX Proteus mirabilis infection; bacterial infection; antibacterial;  
KW Immunosuppressant; gene; da.

XX Proteus mirabilis.

XX US6605709-B1.

XX 12-AUG-2003.

XX 05-APR-2000; 2000US-00543681.

XX 09-APR-1999; 99US-0128706P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton GL;

XX WPI; 2003-895291/82.

XX P-PSDB; ADP05920.

PT New Proteus mirabilis polypeptides and polynucleotides, useful as  
PT reagents for diagnosis of bacterial disease, as components of  
PT antibacterial vaccines, as targets for antibacterial drugs, or as  
PT biocontrol agents for plants.

XX Disclosure; SEQ ID NO 2033; 870bp; English.

CC The invention relates to new Proteus mirabilis polypeptides and  
CC polynucleotides. The invention also relates to antibodies against the  
CC polypeptides, methods for producing the polypeptides, a method of  
CC generating vaccines for immunizing an individual against *P. mirabilis*, a  
CC method for evaluating a compound for the ability to bind a *P. mirabilis*  
CC polypeptide and a method for screening test compounds for anti-bacterial  
CC activity. The polypeptides and polynucleotides are useful as molecular  
CC targets for diagnosing, preventing and treating pathological conditions  
CC resulting from bacterial infection, as reagents for diagnosis of  
CC bacterial diseases, as components of antibacterial vaccines, as targets  
CC for antibacterial drugs or as bio-control agents for plants. This  
CC sequence represents a Proteus mirabilis polynucleotide of the invention.

XX Sequence 726 BP; 224 A; 137 C; 149 G; 216 T; 0 U; 0 Other;

Query Match 48.6%; Score 450.8; DB 10; Length 726;  
Best Local Similarity 76.3%; Pred. No. 5e-139;  
Matches 554; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 92 GTAGCAACATGACAGACCCCGACATTTCTTATCGTTGAAGACGAGTTGTAACAGCAAC 151  
DB 1 GTAGCAAAATATGCAAAACCCCGACATTTCTTATCGTTGAAGATGAAGTACTCTGAAT 60  
QY 152 ACCTGAAAGATATTTTGAAGCGGAGGCTATGATGATGATGATGATGATGATGATGATGAT 211  
DB 61 ACCCTGAAAGATATTTTGAAGCGGAGGCTATGATGATGATGATGATGATGATGATGATGAT 120  
QY 212 GAAATGATCAGATCTCTCTGAATATGATCAACCTGTGATCATGATATCAATCTG 271  
DB 121 GAGATGATCAATATTTCTGTCGACATGATATCAATCTGTGATGATATTAATCTT 180  
QY 272 CCGGTTAGAACGCTTCTGTTAGCGGTGAATCTGCGGACGAGCGAATGTTGGTTG 331  
DB 181 CCGGTTAGAAATGCTTCTTATTTAGCCGCTGAATTAACGTAAGTGAATGATGATTA 240  
QY 332 ATGTTCTGATCGCGGTGACCAAGATGATTAATTTCTGCGCTCGAAATCGGTGCA 391  
DB 241 ATGTTCTGATCAAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 392 GATGATCATGATCAACCAACGTTCAACCGCGTGAATGATGATGATGATGATGATGATGAT 451  
DB 301 GATGATTAATCATCAATTAACATTTAATCTCGTGAATTAATCAATCTGTGATGATTA 360  
QY 452 CTGTCGTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511

Db 361 TTGTCAACGCACTATGATTAATTAGCCAGTGAAGACGATCGTTAGTTGAAAGCTAT 420  
 QY 512 AAGTTCATGTTGGGAACTGACATCAACAGCCGTTGTTGATGCGCCGATGGGAG 571  
 Db 421 AAATTTAATGTTGGAGCTAGATTAATATGTCCTCTTATTAAGCCCTACAGGTAA 480  
 QY 572 CAGTCAACAGCTCCCGGACGAGGATTCGCGCCGATCTTCACTTCTGTAACCCAGGC 631  
 Db 481 CAGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 540  
 QY 632 AAAATTCAGTCCCGGCTGGAATGCTGTAAGAAATGACCGCCGCTGAGCTAAACCGCAC 691  
 Db 541 AAAATTCAGTCCCGGCTGGAATGCTGTAAGAAATGACCGCCGCTGAGCTAAACCGCAT 600  
 QY 692 GACGCTACTGTAAGAGTGAAGATTCGCGCTGTAAGCAATTCGATCAATCAACCGGAT 751  
 Db 601 GATGCTACTGTAAGAGTGAAGATTCGCGCTGTAAGCAATTCGATCAATCAACCGGAT 660  
 QY 752 ACGCCGGAATATCATCGGACCATTCACGCTGTAAGATTCGCTTCCGCTGATCTGAA 811  
 Db 661 ACACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 720  
 QY 812 GATTAA 817  
 Db 721 GAGTGA 726

RESULT 9

AAS92939 ID AAS92939 standard; cDNA; 1048 BP.  
 AAS92939 AC AAS92939;  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #28743.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 OS WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001MO-US008631.  
 PR 31-MAR-2000; 2000US-00540217.  
 PR 23-AUG-2000; 2000US-00649167.  
 XX (HYSB-) HYSBQ INC.  
 PA Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR P-PSDB; ABG28752.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT biotechnology, for genetic disorders or other traits and to assess  
 PT biodiversity.  
 PS Claim 1; SEQ ID NO 28743; 103bp; English.  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a

CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic  
 CC coding sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 1048 BP; 297 A; 265 C; 283 G; 203 T; 0 U; 0 Other;  
 Query Match 48.3%; Score 447.4; DB 5; Length 1048;  
 Best Local Similarity 99.8%; Pred. No. 8.3e-138;  
 Matches 448; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 312 AGCAGGCGAATGTTGCTGATGTTCTGACTGCGCGTGACACGAACTGATAAATTC 371  
 Db 573 AACAGCGCAATGTTGCTGATGTTCTGACTGCGCGTGACACGAACTGATAAATTC 632  
 QY 372 TCGGCTCGAATGCGTGCAGATGATCAATCAACCAACCGTTCAACCGCGTGAATGA 431  
 Db 633 TCGGCTCGAATGCGTGCAGATGATCAATCAACCAACCGTTCAACCGCGTGAATGA 491  
 QY 432 CGATTGCTGACAGCAACCTACTGTCCTGATCATGAACTGAGTACTGTCAGCGAAGAC 491  
 Db 633 CGATTGCTGACAGCAACCTACTGTCCTGATCATGAACTGAGTACTGTCAGCGAAGAC 752  
 QY 492 GTCCGTAAGGTTGAAAGCTCAAGTTCATGTTGGTGAACCTGACATCAACGCGTTGCT 551  
 Db 753 GTCCGTAAGGTTGAAAGCTCAAGTTCATGTTGGTGAACCTGACATCAACGCGTTGCT 611  
 QY 552 TGATCGGCTGATGAGGAGAGTACAAAGCTGCGCGGAGAGGTTCCGGGCAATGCTTC 611  
 Db 813 TGATCGGCTGATGAGGAGAGTACAAAGCTGCGCGGAGAGGTTCCGGGCAATGCTTC 872  
 QY 612 ACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGTAAGAAATGACCG 671  
 Db 873 ACTTCTGTGAAACCCAGGCAAAATTCAGTCCGCTGTAATCTGTAAGAAATGACCG 932  
 QY 672 GCCGTGAGCTGAAACCGGACGACCGTACTGTAAGCGTGAATCCGCGCTGTAATCC 731  
 Db 933 GCCGTGAGCTGAAACCGGACGACCGTACTGTAAGCGTGAATCCGCGCTGTAATCC 992  
 QY 732 ATTTCGAATCTACGCGCGGATACCGCGGAA 760  
 Db 993 ATTTCGAATCTACGCGCGGATACCGCGGAA 1021

RESULT 10

AAT42063\_09/c Continuation (10 of 19) of AAT42063 from base 900001 (Haemophilus influenzae complete ge  
 WP Sequence split into 19 fragments LOCUS AAT42063 Accession Aat42063  
 WP Fragment Name Begin End  
 WP AAT42063\_01 1 110000  
 WP AAT42063\_02 21000 210000  
 WP AAT42063\_03 200001 310000  
 WP AAT42063\_04 300001 410000  
 WP AAT42063\_05 400001 510000  
 WP AAT42063\_06 500001 610000  
 WP AAT42063\_07 600001 710000  
 WP AAT42063\_08 700001 810000  
 WP AAT42063\_09 800001 910000  
 WP AAT42063\_10 900001 1010000  
 WP AAT42063\_11 1000001 1110000  
 WP AAT42063\_12 1100001 1210000  
 WP AAT42063\_13 1200001 1310000  
 WP AAT42063\_14 1300001 1410000  
 WP AAT42063\_15 1400001 1510000

WP AAT42063\_15 1500001 1610000  
WP AAT42063\_16 1600001 1710000  
WP AAT42063\_17 1700001 1810000  
WP AAT42063\_18 1800001 1830121

Query Match 34.5%; Score 319.4; DB 2; Length 110000;

Best Local Similarity 65.8%; Pred. No. 5,4e-94; Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

85 AATTAGTAGCAACATGACAGACCCGACATTTCTATGTTGAAGACGATTGGTAAC 144  
Db AGATAGAGATGGAATAATGACTCTCCAAAATTTCTGTTGAAAGATATAATTTGTAC 36581  
145 ACGCAACACGTTGAAAAGTATTTTGAAGCGAAGCTATGATGTTTTCGACGACAGA 204  
Db TCGAATATACGCTTAAGGGATTTTGAAGCGAAGATGATGTTTGAACAGAAA 36521  
205 TGGCGGGAATATCATCATCTCTCTGTAATATGATCATACCTGTGATCATGATAT 264  
Db TGGTGTGAATGATCATCATATATTTGCAATCATATATATTTGGTGTGATGATAT 36461  
265 CATCTGCGGGTAAGACGGTCTTCTGTTAGCGCGAATCTGGAGAGCGGATGT 324  
Db TAATTTACAGGCAAAACGGCTTATTTATTTGCAAGAACTCCGTGAAGATTAAGCTT 36401  
325 TGGTGTGATGTTCTCTGATGCGCGTGAACAAGATGATTAATTTCTCGGCTCGAAAT 384  
Db ACCCTTATTTTATTTTAACTGCTGAGATATGAAAGGATTAATTTTGGTTGGAAT 36341  
385 CGGTGAGATGACTATCATCAACCAACCGTTCAACCGCGTGAAGTCAATGTTGTGACG 444  
Db TGGTCCGATGACTATTTGCTAAGCCTTTTAACTGAGAACTGATCTCCGCGACG 36281  
445 CAACCTACTGTCGCGTACCAATGAGTCTGACGAGAAAGTGTGACGTTGA 504  
Db TAATCTATTGATGCTGCAATG---CCGATCAAGAAAAGAAATATCATTTGTGAGAG 36224  
505 AAGCTCAAGTTCAATGTTTGGAGACTGACATCAACGCGCTTCTGTTGATCGGCTTGA 564  
Db ATTCTATCGTTTAAATGTTTGAATATTAAGCTTAATTAAGCAGATTAATTAACACAG 36164  
565 TGGCGACAGTACAGCTGCGCGACGAGATTTCCGCGCATCTTCACTTCTGTGAAA 624  
Db AGGACACAGAAATTCAACTTCTGCGAGTGAATTTCTGCAATGTTACATTTCTGTGAAA 36104  
625 CCAGGCAAAATTCAGTCCCGTGTAACTGAGAAATTAACCGCGCTGAGCTGA 684  
Db TCCAGAAAATTCGCAACGCGCAAGATTTGAAAAAATTAACCGAGAGATTA 36044  
36103 TCCAGAAAATTCGCAACGCGCAAGATTTGAAAAAATTAACCGAGAGATTA 36044  
685 ACCGACAGACCGTACTGATGACGTGACATCCGCGTATTCGTAACATTTGCAATCTAC 744  
Db ACCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 36043  
745 GCGGATACCGCGGAATATCAGCCACATTCAGCGTGAAGTATGCTTGTGCGGTGA 804  
Db TCCCAATACCTCCAAATATCATTTATGACTATACATGAGAAAGCTATGTTTGTGAGAG 35983  
805 TCTGGAAGA 813  
Db TATGTAATA 35915

RESULT 11  
AAS53409

ID AAS53409 standard; DNA; 711 BP.

AC AAS53409;

DT 13-FEB-2002 (first entry)

DE Haemophilus influenzae DNA for cellular proliferation protein #191.

XX Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;

KW antibacterial; drug design.  
OS Haemophilus influenzae.  
XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US009180.

XX 21-MAR-2000; 2000US-0191078P.

XX 23-MAY-2000; 2000US-0206848P.

XX 26-MAY-2000; 2000US-0207727P.

XX 23-OCT-2000; 2000US-0242578P.

XX 27-NOV-2000; 2000US-0253625P.

XX 22-DEC-2000; 2000US-0257931P.

XX 16-FEB-2001; 2001US-0269308P.

XX (BLIT-) ELITRA PHARM INC.

XX Haseelbeck R, Ohlsen KL, Zyskind JM, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

XX P-PSDB; A035550.

XX New polynucleotides for the identification and development of

XX antibiotics, comprise sequences of antisense nucleic acids.

XX Claim 27; SEQ ID NO 7046; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the genes,  
XX their use in the discovery of novel antibiotics, the essential genes,  
XX themselves and the encoded proteins. The prokaryotes used are Escherichia  
XX coli, Staphylococcus aureus, Enterococcus faecalis. The invention is also  
XX useful for the identification of potential new targets for antibiotic  
XX development. The antisense nucleic acids can also be used to identify  
XX proteins used in proliferation, to express these proteins, and to obtain  
XX antibodies capable of binding to the expressed proteins. The proteins can  
XX be used to screen compounds in rational drug discovery programmes. The  
XX antisense nucleic acid sequence is also useful to screen for homologous  
XX CC nucleic acids which are required for cell proliferation in a wide variety  
XX of organisms. The present sequence encodes an essential prokaryotic  
XX cellular proliferation protein. Note: The sequence data for this patent  
XX did not form part of the printed specification, but was obtained in  
XX electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 711 BP; 246 A; 113 C; 143 G; 209 T; 0 U; 0 Other;

XX Query Match 34.1%; Score 316.2; DB 4; Length 711;

XX Best Local Similarity 66.2%; Pred. No. 3,6e-94; Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

XX 101 ATGCAAGACCCGACATTTCTTATCGTTGAAGCGAGTTGTAAACAGCAACGTTGAAA 160

XX 1 ATGACTACTCCAAAATTTCTGTTGAAGATGAATTTGATCTGAAATATAGCTTTAA 60

XX 161 AGTATTTGCAAGCGAAGCTATGATGTTTTCGAAGCGACAGATGCGCGGAATGCAT 220

XX 61 GGGATTTTGAAGCGAAGATATGATGTTTGAAGCGAAGATGTTTGAAGATGAT 120

XX 221 CAGATCTCTGATATATGATCAACCTGGTATGATGATATCAATCTGCCGGTAA 280

XX 121 CATATTTGCAAAATCATATATATTTATTTGTTGATGATATTAATTTACAGGCAAA 180

XX 281 AACGCTCTGTTTGGCGGCGTGAACGCGGAGCGGCAAGTGGCGTTGATGTTCTG 340

XX 181 AACGCTTATTTTGGCAAGAACTCGGTGAAGATTAAGCTTACTCTTTATTTTAA 240

XX 341 ACTGCGGTGACAAAGACTGATTAATTTCTGCGCTCGAAATCGGTGACATGACTAC 400

Db 241 ACTGTCGAGATTAATGAGTGAATTAATTTAGTTGGAATTTGTCGCAATGACTAT 300  
 Qy 401 ATCAACCAACCGTTCAACCGCGGGAACGATTCGTCGACGCAACCTACTGTCCT 460  
 Db 301 TTGACTAAGCCTTTTAACCTTAAGAACTGACTATCCGGCAGTAACTTATGATCGT 360  
 Qy 461 ACCATGAATTTGGGTACTGTCAAGCAAGAACTGTCGAGTGAAGCTACAAAGTTCAT 520  
 Db 361 GCATG---CCGCATCAAGAAAAAGAAATTAATTTGTCGAGAAATTTCTATGTTTAT 417  
 Qy 521 GATTGGAACTGGAATCAACACCGCTTGTGATGCGCTGATGCGAGCACTACAAG 580  
 Db 418 GATTGAAATTAAGCTTAATTAATTAACCAAGTTTAATTAACCAAGAAAGCAAGAAATTCAA 477  
 Qy 581 CTGCGCGGAGCGAGTTCCGCGCATCTTCACTTCTGGAAGAAACCGGAAATTTGAG 640  
 Db 478 CTTCCTCGAGTGAATTTGTCGCAATTTCTGGAATTTCCGAAATTTGCA 537  
 Qy 641 TCCCGTCTGAACTGCTGAAGAAATTAACCGCGCTGAGCTGAAACCGCAAGACCTACT 700  
 Db 538 AGCGCGGAGAAATTTGCTGAAGAAATTAACCGCGAGAGTTAAACCTCAAGATCGTACC 597  
 Qy 701 GTAGACGTGACGATCCGCGTATTTGTTAACTTTTCAATCTAGCCGAGTACCGCGGAA 760  
 Db 598 GTAGATGTCAATTCGAGATTCGAAATTTTGAAGATCAATCCCAATATCCCAAT 657  
 Qy 761 ATCATGCGCACCATTTCAAGGTTATGCTTCCGCGTGAATCTGGAAGA 813  
 Db 658 ATCATTAATGACTAATACATGAGAAAGGCTATGTTTGTGGAATATTTGATA 710

RESULT 12  
 ID ACA34287  
 AC ACA34287 standard; DNA; 711 BP.  
 XX  
 AC ACA34287;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Prokaryotic essential gene #15944.  
 XX  
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;  
 FW drug design; gene.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 MO200277183-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 21-MAR-2002; 2002WO-US009107.  
 XX  
 PR 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX  
 PA (ELIT-) ELITRA PHARM INC.  
 XX  
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 DR P-PsDB; ABU30417.  
 XX  
 PT New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 PS  
 PS Claim 14; SEQ ID NO 22157; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC the gene product or that has an activity against a biological pathway  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs; or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target  
 CC prokaryotic essential genes. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

Sequence 711 BP; 246 A; 113 C; 143 G; 209 T; 0 U; 0 Other;

Query Match 34.1%; Score 316.2; DB 8; Length 711;  
 Best Local Similarity 66.2%; Pred. No. 3.6e-94;  
 Matches 472; Conservative 0; Mismatches 238; Indels 3; Gaps 1;

Qy 101 ATGACAGACCCCGACATTTCTTATCGTTGAAGACGAGTTGTAACAGCAACAGTTGAA 160  
 Db 1 ATGACTATCTCAAAATTCGTTGTTGAAGTGAATTTGTCATCGAAATGACGTTAA 60  
 Qy 161 AGTATTTTCAGACCGAGAGGCTATGATGTTTTCGAAGCAGATGCGCGGAATGCAT 220  
 Db 61 GGGATTTTTCAGACCGAGAGATATGATGTTTTCGAAGCAGAAATGTTGAATATGCAT 120  
 Qy 221 CAGATCTCTCTGAATATGATCAATCACTGTGATCATGATATCATCTGCGGGTAA 280  
 Db 121 CATATATGCAAAATCATATATTAATTTGTTGATGATGATTAATTTACAGGCAAA 180  
 Qy 281 AACGCTCTTCTGTAGCGGCTGAACGCGCGAGCAGAGGAATGTTGCTTATGTTCTG 340  
 Db 181 AACGCTTATTTATGTCAGAGAACTCCGTGAAGATTAACCTTATTTTATTTTAA 240  
 Qy 341 ACTGCGCGTGAACAGAACTGATTAATTTCTGCGCTCGAATCGGTGAGATGATAC 400  
 Db 241 ACTGTCGAGTAATGAAGTGAATTTTAAAGTTTGAAGTTGTCGCGGATGATCAT 300  
 Qy 401 ATCAACCAACCGTTCAACCGCGGGAACGATTCGTCGACGCAACCTACTGTCCTG 460  
 Db 301 TTGACTAAGCCTTTTAACCTTAAGAACTGACTATCCGGCAGTAACTTATGATCGT 360  
 Qy 461 ACCATGAATTTGGGTACTGTCAAGCAAGAACTGTCGAGTGAAGCTACAAAGTTCAT 520  
 Db 361 GCATG---CCGCATCAAGAAAAAGAAATTAATTTGTCGAGAAATTTCTATGTTTAT 417  
 Qy 521 GATTGGAACTGGAATCAACACCGCTTGTGATGCGCTGATGCGAGCACTACAAG 580  
 Db 418 GATTGAAATTAAGCTTAATTAATTAACCAAGTTTAATTAACCAAGAAAGCAAGAAATTCAA 477  
 Qy 581 CTGCGCGGAGCGAGTTCCGCGCATCTTCACTTCTGGAAGAAACCGGAAATTTGAG 640  
 Db 478 CTTCCTCGAGTGAATTTGTCGCAATTTCTGGAATTTCCGAAATTTGCA 537



QY 641 TCCCGTCTGAAGTCTGAAGAAATGACCGCGGTGAGTGAACCGCAGACCGTACT 700  
DB 538 ACCCGGGAAGAAATGCTGAAAATAATGACCGGACGAGATTAATAACTCAGGATCGTACC 597  
QY 701 GTGAGCTGACGATCCGCCGTATTCGTAACATTTCGATTCAGCCGGATAGCCGGAA 760  
DB 598 GTGAGTGTCAAAATTCGACCTATCAGAAAACATTTTGAGATCATCCCAATCTCCAAAT 657  
QY 761 ATCATGCGCCACATTCGCGGTGAGTATCGCTTCGCGGTGATCTGAGAA 813  
DB 658 ATCATTTATGACTTATCATGAGAAAGCTATCGTTTGTGAGATTTGAATA 710

RESULT 13  
AAS82422  
ID AAS82422 standard; cDNA; 1110 BP.  
XX AAS82422;  
AC AAS82422;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #18226.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
PI Dmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG18235.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 1; SEQ ID NO 18226; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention. Note: The sequence data for this patent did not appear in the invention. Note: The sequence data for this electronic format directly from WIPO at [ftp://ipo.int/pub/published\\_pct\\_sequences](ftp://ipo.int/pub/published_pct_sequences)

XX SQ Sequence 1110 BP; 256 A; 291 C; 296 G; 267 T; 0 U; 0 Other;  
Query Match 25.1%; Score 232.8; DB 5; Length 1110;  
Best Local Similarity 98.8%; Pred. No. 2.9e-66;  
Matches 245; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 544 CCGTGTGATGATCGGCGCTGATGCGAGCAGTA-CAAGCTGCCGCGCAGGATTCGCG 602  
DB 761 CGTTGGTTGATATGGCGCTGATGGCGAGCGATCAAGCTGCCGCGAGGATTCGCG 820  
QY 603 CCATGCTTCACTTCTGTGAATAACCCAGCAAAATTGATCCGTCTGAACTGCTGAGA 662  
DB 821 CCATGCTTCACTTCTGTGAATAACCCAGCAAAATTGATCCGTCTGAACTGCTGAGA 880  
QY 663 AAATGACCGGCGGTGAGCTGAACCCGACGACCGTACTGAGACGTGAGATCCGCGCTA 722  
DB 881 AAATGACCGGCGGTGAGCTGAACCCGACGACCGTACTGAGACGTGAGATCCGCGCTA 940  
QY 723 TTGTAACATTTGATCTGATACCGCGGATACCGCGGAAATCATCGCCATTCAGCGTG 782  
DB 941 TTGTAACATTTGATCTGATACCGCGGATACCGCGGAAATCATCGCCATTCAGCGTG 1000  
QY 783 AAGCTTAT 790  
DB 1001 AAGTTCAT 1008

RESULT 14  
ABQ90117  
ID ABQ90117 standard; DNA; 729 BP.  
XX  
XX ABQ90117;  
AC  
XX  
DT 01-OCT-2002 (first entry)  
XX  
DE M. capsulatus gene #102 for DNA array.  
XX  
KW Micro array; gene; ds; differential expression; gene expression.  
XX  
OS Methylococcus capsulatus.  
XX  
PN WO200255655-A2.  
XX  
PD 18-JUL-2002.  
XX  
PF 14-JAN-2002; 2002WO-NO000019.  
XX  
PR 12-JAN-2001; 2001NO-00000235.  
PR 12-JAN-2001; 2001NO-00000239.  
XX  
XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
PA (TIGR-) TIGR.  
XX  
XX Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;  
PI Lillhaug JR, Lossius I, Bisen UB, Fraser CM, Durkin AS;  
PI Salzberg SL;  
XX  
DR WPI; 2002-557818/59.  
XX  
XX Novel DNA array useful for determining differential expression of  
PT Methylococcus capsulatus genes, comprises polynucleotides or  
PT oligonucleotides representative for a selective number of Methylococcus  
PT capsulatus genes.  
XX  
XX Claim 19; Page 99; 678bp; English.

The invention relates to a novel DNA array giving a representation of a number of Methylococcus capsulatus genes. The method of the invention is useful for determination of the differential expression of the genes of M. capsulatus, and for studying gene expression on a genomic scale and in gene expression assays of M. capsulatus genes. The sequences shown in ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the

CC invention  
 XX Sequence 729 BP; 136 A; 241 C; 222 G; 130 T; 0 U; 0 Other;  
 SQ  
 Query Match 14.8%; Score 137.4; DB 6; Length 729;  
 Best Local Similarity 49.8%; Pred. No. 1.4e-34;  
 Matches 348; Conservative 0; Mismatches 351; Indels 0; Gaps 0;  
 QY 104 CAGACCCCGGCAATTTCTTATGCTTGAAGACGATGTTGAACACGCAACAGTTGAAAGT 163  
 DB 10 CCGACCAACAGATATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 69  
 QY 164 ATTTTCAAGCGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 223  
 DB 70 TACCTTTCGCGCTTTCGCGATGGAAGCCGCTCGTTCATGATGATGATGATGATGATGATGAT 129  
 QY 224 ATCTCTCTGAAATATGATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 283  
 DB 130 GCGCTGACGGAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 189  
 QY 284 GGTCTTCTGTTTGAAGCGGCTGAACCTGCGGACGACGACGACGACGACGACGACGACGAC 343  
 DB 190 GGCCTGACCCCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 249  
 QY 344 GGCCTGACCAACGAAGTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 403  
 DB 250 GCGACGGGCGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309  
 QY 404 ACCAACCCTTTCACACCCGCTGAACCTGATGATGATGATGATGATGATGATGATGATGAT 463  
 DB 310 GCCAAGCGCTTTCGACCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369  
 QY 464 ATGATTTGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 523  
 DB 370 CCGACGACGAGCGCAACCGCAACCGCAACCGCAACCGCAACCGCAACCGCAACCGCAAC 429  
 QY 524 TGGGAATGACATCAACACCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 583  
 DB 430 TGGGCGGTGACCGGGGTGCTCCGGCACTGATCTCGCCGCAAGGCAATGCTGCTCTCTTC 489  
 QY 584 CCGCGGACGAGTTCGCGCGCATGCTTCACTTCTGTGAAACCCAGCAAAATTCAGTCC 643  
 DB 490 TCCAAATGCCAGATTCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 549  
 QY 644 CGTGTGAACCTGCTGAAAGAAATGACCGGCGCTGATGATGATGATGATGATGATGATGATGAT 703  
 DB 550 CCGGATCAGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 609  
 QY 704 GACGTGACGATCCGCGCTATTCGTAACATTCGATGATGATGATGATGATGATGATGATGAT 763  
 DB 610 GACCTGCTGATTCGCGGCTGCGGCGCAAGAACTGCAAGATGATGATGATGATGATGATGAT 669  
 QY 764 ATGCGCACATTCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 802  
 DB 670 ATCAAGACCATCGCGCGAAGTTATTTCTTCAAGCGCT 708  
 RESULT 15  
 ABD17813  
 ID ABD17813 standard; DNA; 837 BP.  
 XX  
 AC ABD17813;  
 XX  
 DT 29-JUL-2004 (first entry)  
 XX  
 DE Pseudomonas aeruginosa polynucleotide #16417.  
 XX  
 KW Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;  
 XX antibacterial.  
 OS Pseudomonas aeruginosa.  
 XX  
 PN US551795-B1.

XX  
 PD 22-APR-2003.  
 XX  
 PF 18-FEB-1999; 99US-00252991.  
 XX  
 PR 18-FEB-1998; 98US-0074788P.  
 PR 27-JUL-1998; 98US-0094190P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Rubenfield MJ, Nolling J, DeLoughery C, Bush D;  
 XX  
 DR WPI; 2003-615309/58.  
 XX  
 DR P-PDSB; ABO84242.  
 XX  
 PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,  
 PT useful as molecular targets for diagnostics, prophylaxis and treatment of  
 PT pathological conditions resulting from bacterial infection.  
 XX  
 PS Disclosure; SEQ ID NO 16417; 455bp; English.  
 XX  
 CC The invention relates to Pseudomonas aeruginosa polypeptides and the  
 CC polynucleotides encoding them. The sequences are useful in diagnosis and  
 CC therapy of pathological conditions, as molecular targets for diagnostics,  
 CC prophylaxis and treatment of pathological conditions resulting from a  
 CC bacterial infection, for evaluating a compound, such as a polypeptide,  
 CC for the ability to bind a P. aeruginosa nucleic acid, as components of  
 CC effective antibacterial targets, as targets for antibacterial drugs,  
 CC including anti-P. aeruginosa drugs, as templates for recombinant  
 CC production of P. aeruginosa-derived peptides or polypeptides, as target  
 CC components for diagnosis and/or treatment of P. aeruginosa-caused  
 CC infection, and in detection of P. aeruginosa sequences or other sequences  
 CC of Pseudomonas species using biochip technology. Sequences ABD1397-  
 CC ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification and was obtained in electronic format from USPTO at  
 CC seqdata.uspto.gov/sequence.html  
 XX  
 SQ Sequence 837 BP; 140 A; 303 C; 273 G; 121 T; 0 U; 0 Other;  
 Query Match 13.6%; Score 126; DB 11; Length 837;  
 Best Local Similarity 52.9%; Pred. No. 9.7e-31;  
 Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;  
 QY 243 TCAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302  
 DB 263 TCGAGCTGTGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 322  
 QY 303 AACTGCGGACGAGCGCAATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 362  
 DB 323 AGCTGCGGCTGCTCGGAGGTGGGATCACTGATGATGATGATGATGATGATGATGATGATGAT 382  
 QY 363 ATAAATTTCTGCGCTGGAATGCGTGCATGATGATGATGATGATGATGATGATGATGATGATGAT 422  
 DB 383 ATGCAATGCTGCGCTGGAATGCGGCGGCAATGATGATGATGATGATGATGATGATGATGATGAT 442  
 QY 423 GTGAATGACGATTCGTCACGCAACCTACTGCTCCGCTACATGAAATCTGGTACTGTCA 482  
 DB 443 GCGAATGCTGTGTGCGGGG---GAAGAATGATATCCCGGGGTGCCCATGCCAGGCCA 499  
 QY 483 GCGAAGAACGTGATGCTGTTGAAAGCTCAAGTTCAATGATGATGATGATGATGATGATGATGAT 542  
 DB 500 GCGCGGCGCCCGCGCGGACGCTCAAGGATGCTGCGGCTGCTGATGATGATGATGATGATGAT 559  
 QY 543 GCGCTTGTGATGCGCGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 602  
 DB 560 GCGCGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 619  
 QY 603 CAATGCTTCACTTCTGTAAGAAACCGGCAAAATTCAGTCCGCTGATGATGATGATGATGATGAT 662  
 DB 620 TGCTCGGCGCTTTCGCGGCAACAGGCGGATATCCCTGAGCGCGGACGATGATGATGATGATGAT 679  
 QY 663 AAATGACCGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 722

Db	680	AGATCCGCAACCGGAGTGGCTGCCCCAGCATCGTCCATCGAGTGTGTCGGCCG	739
Qy	723	TTGGTAAACATTTGAACTACGCGGATACCGGAAATCATGCGCCACCATTCACGGTG	782
Db	740	TGGTGGCAAGCTGCGGAGCAGCCCGGCCGAACCGCAACTGATCATCACATCCACGGCG	799
Qy	783	AAGTTATCGCTTCTGCG	800
Db	800	CCGGTACCTGTTCACCG	817

Search completed: January 22, 2005, 12:42:09  
 Job time : 556 secs

**This Page Blank (uspto)**



Db 361 CTCCTCCGTAACATGAACTCCGTAACCTGTAGACGAGACCGTCCGAGCGGTGGAAGATAC 420

QY 512 AAGTTCAATGTTGGGAATGACATCAACAGCCGTTGTTGATGCGCCCTGATGCGAG 571

Db 421 AAGTTCAACGGCTGGGAATCTGATATCAACAGCCGTTCCCTGTTAGCCCGAAGCGGAA 480

QY 572 CAGTCAAGCTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 631

Db 481 CAGTCAAGCTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 540

QY 632 AAAATTCAGTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 691

Db 541 AAAATTCAGTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 600

QY 692 GACCTGACTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 751

Db 601 GACCTGACTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

QY 752 ACGCCGAAATCATGCGACGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 811

Db 661 ACGCCGAAATCATGCGACGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 720

QY 812 GATTAA 817

Db 721 GAATTA 726

RESULT 2

US-09-543-681A-2033

Sequence 2033, Application US/09543681A

Patent No. 6605709

GENERAL INFORMATION:

APPLICANT: GARY BRETTON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS

FILE REFERENCE: 2709.1002-001

CURRENT APPLICATION NUMBER: US/09/543,681A

PRIOR FILING DATE: 2000-04-05

PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 8344

SEQ ID NO 2033

LENGTH: 726

TYPE: DNA

ORGANISM: Proteus mirabilis

US-09-543-681A-2033

Query Match 48.6%; Score 450.8; DB 4; Length 726;

Best Local Similarity 76.3%; Pred. No. 5e-152;

Matches 554; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY 92 GTAGCAAAACATGAGACCCGACATTCCTTATGTTGAGAGAGAGTGTGTAACAGCAAC 151

Db 1 GTAGCAAAATATGAAACCCGACATTCCTTATGTTGAGAGAGAGTGTGTAACAGCAAC 60

QY 152 ACCTGAAAGATATTTTGAAGCGAAGCTATGATGTTTGAAGCGACAGATGAGCGG 211

Db 61 ACCCTGAAAGATATTTTGAAGCGAAGCTATGATGTTTGAAGCGACAGATGAGCGG 120

QY 212 GAAATGATCAGATCTCTCTGTAATATGATCAACCTGGTGATGATGATATCAATCTG 271

Db 121 GAAATGATCAGATCTCTCTGTAATATGATCAACCTGGTGATGATGATATCAATCTG 180

QY 272 CCGGGTAAGAACGATCTCTCTGTAATATGATCAACCTGGTGATGATGATATCAATCTG 331

Db 181 CCGGGTAAGAACGATCTCTCTGTAATATGATCAACCTGGTGATGATGATATCAATCTG 240

QY 332 AAGTTCTGACTGCGCGGTGACAGAAAGTGTGTAATATTCGCGCTGAAATCGGTGCA 391

Db 241 AAGTTCTGACTGCGCGGTGACAGAAAGTGTGTAATATTCGCGCTGAAATCGGTGCA 300

QY 392 GATGACTACATCAACCAACCGTTGAAACCGGTGAAACCGGTGAAACCGGTGAAACCGT 451

Db 301 GATGATTACTACTAAACCAATTAAATCTCTGTAATTAATATCCGTCCTGTAATCTTA 360

QY 452 CTGTCCGTAACATGATCTGAGTACTGTCAGGAGAAAGTGTGAGCGTTGAAAGCTAC 511

Db 361 TTGTACAGCACTATGATTTAGAGATGACAGAAAGACGTCGTTAGTTGAAAGCTAT 420

QY 512 AAGTTCAATGTTGGGAATCTGACATCAACAGCCGTTGATGATGATGATGATGATGAT 571

Db 421 AAGTTCAATGTTGGGAATCTGACATCAACAGCCGTTGATGATGATGATGATGATGAT 480

QY 572 CAGTCAAGCTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 631

Db 481 CAGTCAAGCTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 540

QY 632 AAAATTCAGTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 691

Db 541 AAAATTCAGTCCCGGACGAGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 600

QY 692 GACCTGACTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 751

Db 601 GACCTGACTGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660

QY 752 ACGCCGAAATCATGCGACGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 811

Db 661 ACGCCGAAATCATGCGACGATTCGCGCCATGCTTCACTTCTGTGAAAACCCAGGC 720

QY 812 GATTAA 817

Db 721 GATGTA 726

RESULT 3

US-09-557-884-1/c

Sequence 1, Application US/09557884

Patent No. 6506581

GENERAL INFORMATION:

APPLICANT: Fleischmann et al.

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/557,884

FILING DATE: 25-Apr-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,102

FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB186F3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-309-8504

TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;  
Best Local Similarity 65.8%; Pred. No. 4.4e-102;  
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

85 AATTAGTAGCAACATGACAGACCCGACATCTTATCGTTGAAGACGATGTTGTAAC 144  
Db 936640 AGATATAGATGGAATATGCTCTCCAAAATTTCTGTTGGAAGATGAATTTGCAC 936581  
Qy 145 ACCGAACAGCTTGAAGATTTTTCGAGCGGAAGCTATGATGTTTTCGAACGACAGA 204  
Db 936580 TCCAAATACGCTTAAAGGGATTTTTCGAGCGGAAGATGATGTTTTCGAACGAGAAA 936521  
Qy 205 TGGCGGGAATGATCATGATCTCTCTGTAATGATCATGACCTGGTATGATGATAT 264  
Db 936520 TGGTGTGAATGATCATGATATGATGGAATCATTAATTAATTTGGTGTGATGATAT 936461  
Qy 265 CAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCGCGTGAACCTGCGACGACGAGATGT 324  
Db 936460 TAATTTACAGGCAAAAAGCGCTTATTTATGGCAAGAACTCCGTGAAGAATTAAGCTT 936401  
Qy 325 TGGCTGATGTTCTCTGATCGCGCTGACAAAGAGTGAATTAATTTCTCGGCTCGAAT 384  
Db 936400 ACCCTTAATTTTAACTGCTGAGATTAATGAAGTGAATTAATTTAGTTTGGAAAT 936341  
Qy 385 CGGTGAGATGATACATACCAACCGTTCAACCGCGGTGAACCTGACGATGCTGACAG 444  
Db 936340 TGGTGCAGATGATATTTTGAAGCTTTTAACTTGAAGAACTGATGATGCGGACAG 936281  
Qy 445 CAACCTACTGTCCTGATACATGATGAGTCTGATGACGGAAGAACTGATGAGCTTGA 504  
Db 936280 TAATCTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936224  
Qy 505 AAGCTCAAGTTCAATGTTGGGAATGACATCAACAGCGCTTCTGATGATGCGCTCTGA 564  
Db 936223 ATCTATCTGTTTAAAGTTGGAATTAAGACCTAAATAGCAAGTTTAAATTAACACAGA 936164  
Qy 565 TGGCGAGATGATACAGTGGCGGCGGACGAGTTCGCGGCACTGATCTGATGATAA 624  
Db 936163 AGGACAGAAATTAATTAATCTCTGCGAGTGAATTTCTGCAATGTTATCTGTTGAAA 936104  
Qy 625 CCCAGGCAAAATTCAGTCCGCTGCTGAATGCTGAAGAAATGAACGCGCGTGAAGCTGA 684  
Db 936103 TCCAGGAAATTTCCAAACGCGCAAGATTTCTGAAATTAATGACCGGAGAGAGTTAAA 936044  
Qy 685 ACCGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 744  
Db 936043 ACCTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 935984  
Qy 745 GCGGATACCGCGGAATATTCGCGACATTCAGCGGGAAGTTATGCTTTCGCGGTGA 804  
Db 935983 TCCCAATATCTCAAAATATATATGATGATGATGATGATGATGATGATGATGATGAT 935924  
Qy 805 TCTGAAGA 813  
Db 935923 TATTGAATA 935915

## RESULT 4

US-09-643-990A-1/c  
Sequence 1, Application US/09643990A  
Patent No. 6528289

GENERAL INFORMATION:  
APPLICANT: Robert D. Fleischmann

Mark D. Adams  
Owen White  
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of  
the Haemophilus influenzae Rd Genome, Fragments  
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville,  
STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium  
OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07  
APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:  
NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302  
REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1830121 base pairs

TYPE: nucleic acid  
STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-643-990A-1

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;  
Best Local Similarity 65.8%; Pred. No. 4.4e-102;  
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

85 AATTAGTAGCAACATGACAGACCCGACATCTTATCGTTGAAGACGATGTTGTAAC 144  
Db 936640 AGATATAGATGGAATATGCTCTCCAAAATTTCTGTTGGAAGATGAATTTGCAC 936581  
Qy 145 ACCGAACAGCTTGAAGATTTTTCGAGCGGAAGCTATGATGTTTTCGAACGACAGA 204  
Db 936580 TCCAAATACGCTTAAAGGGATTTTTCGAGCGGAAGATGATGTTTTCGAACGAGAAA 936521  
Qy 205 TGGCGGGAATGATCATGATCTCTCTGTAATGATCATGACCTGGTATGATGATAT 264  
Db 936520 TGGTGTGAATGATCATGATATGATGGAATCATTAATTAATTTGGTGTGATGATAT 936461  
Qy 265 CAATCTGCGGGTAGAAGACGGTCTTCTGTTAGCGCGTGAACCTGCGACGACGAGATGT 324  
Db 936460 TAATTTACAGGCAAAAAGCGCTTATTTATGGCAAGAACTCCGTGAAGAATTAAGCTT 936401  
Qy 325 TGGCTGATGTTCTCTGATCGCGCTGACAAAGAGTGAATTAATTTCTCGGCTCGAAT 384  
Db 936400 ACCCTTAATTTTAACTGCTGAGATTAATGAAGTGAATTAATTTAGTTTGGAAAT 936341  
Qy 385 CGGTGAGATGATACATACCAACCGTTCAACCGCGGTGAACCTGACGATGCTGACAG 444  
Db 936340 TGGTGCAGATGATATTTTGAAGCTTTTAACTTGAAGAACTGATGATGCGGACAG 936281  
Qy 445 CAACCTACTGTCCTGATACATGATGAGTCTGATGACGGAAGAACTGATGAGCTTGA 504  
Db 936280 TAATCTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936224  
Qy 505 AAGCTCAAGTTCAATGTTGGGAATGACATCAACAGCGCTTCTGATGATGCGCTCTGA 564  
Db 936223 ATCTATCTGTTTAAAGTTGGAATTAAGACCTAAATAGCAAGTTTAAATTAACACAGA 936164

```
QY 565 TGGCGAGCAGTACAGCTGCCGCGCAGAGTTCCGCGCATGCTTCTGTGAAA 624
Db 936163 AGGACAGAAATTCAACTTCTCTGCATGTAATTCGCAATGTTACATTTCTGTAAAA 936104
QY 625 CCCAGGCAAAATTCAGTCCCGTGTGTACTGTCTGAAAGAAATGACCGCGCTGAGCTGAA 684
Db 936103 TCCAGGAAATTTGCAAAAGCGCGCAAGAAATTTGTGAAAAAATGACCGGAGAGATTAA 936044
QY 685 ACCGCGACGACGTAAGTGTAGACGTGACGATCCGCGGTAATTCGTAACATTTGGAATTCAC 744
Db 936043 ACCTCAGAGATCGTACCGTATAGTCAATTCGACCTATCAGAAAACATTTTGAAGATCA 935984
QY 745 GCGGATAGCGCCGGAATCATGCCACCAATTCACGGTGAAGTTATCGCTTCTGCGGTGA 804
Db 935983 TCCCAATACTCCAAATATCATTAATGACTATATCATGAGGAAGGCTATCGTTTGTGGAGA 935924
QY 805 TCTGGAAGA 813
Db 935923 TATTGAATA 935915
```

```
RESULT 5
US-10-329-960-1/c
; Sequence 1, Application US/10329960
; Patent No. 6742927
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; Patent No. 6742927
; FILE REFERENCE: Thereof, and Uses Thereof
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIOR FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
```

```
OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
```



```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248) .. (107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136) .. (117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750) .. (119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924) .. (119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038) .. (120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344) .. (121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167) .. (122167)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336) .. (122336)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340) .. (131340)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360) .. (131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910) .. (139910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398) .. (140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750) .. (142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058) .. (145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145171) .. (145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145942) .. (145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197) .. (147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (150841) .. (150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:

```

```

NAME/KEY: misc_feature
LOCATION: (152500) .. (152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature

```

```

Query Match 34.5%; Score 319.4; DB 4; Length 1830121;
Best Local Similarity 65.8%; Pred. No. 4,4e-102;
Matches 480; Conservative 0; Mismatches 246; Indels 3; Gaps 1;

```

```

QY 85 AATTAGATGAGAAATGAGACCCCGCATCTTATCGTTGAAGAGAGTGGTAAC 144
DB 936640 AGTATAGATGAGAAATGAGACTCTCAAAAATTCGTTGTAAGATGAAAATGTGCAC 936581
QY 145 ACCGAACAGCTTGAAGATTTTTCGAAGCGAAGGCTATGATGTTTTCGAACGACAGA 204
DB 936580 TCGAAATACGCTTAAAGGATTTTTCGAAGCGAAGGATATGATGTTTTCGAACGACAAA 936521
QY 205 TGGCGCGGAATGATCATGATCCTCTCTGAAATGACATCAACCTGATCATGATAT 264
DB 936520 TGGTGTGAAATGATCATGATATTTGCAAAATCATATATTAATTTGTTGATGATAT 936461
QY 265 CAATTCGCGGGTAAAGACGCTTCTGTTAGGCGGTGAACCTGCGGAGCGCAATGT 324
DB 936460 TAATTTACGAGGAAAGCGCTTATTTATTTGCAAGAACTCTGGAAGAAATTAAGCTT 936401
QY 325 TGGCTGATGTTCTCTGATGCGCGGTGACAAAGATCGATTAATTTCTCGGCTCGAAT 384
DB 936400 ACCTCTATTTTATTTTAACTGCTGATGATTAAGATGATTAATTTTATGATTTGAAAT 936341
QY 385 CGGTGAGATGACTATCATCAACCAACCGTTCAACCGCGGTGAACGATCGTGCACG 444
DB 936340 TGGTGCAGATGACTATTTGATCTAAGCCTTTTAACTGAGAACTGATCTGCGGACG 936281
QY 445 CAACCTACGTCCTCCGATCACTGATCTGGTACTGTCAGAGAAAGACGTGACGTGA 504
DB 936280 TAATCTATTTGATCTGCAATG--CCGATCAAGAAAGAAATTAATTTGTCGAGA 936224
QY 505 AAGCTCAAGTTCAATGATGTTGGAACTGACATCAACAGCCGTTGATCGGCCCTGA 564
DB 936223 ATTCTATCGTTTAAATGTTGGAATTAACCTTAATTAAGCACAGTTTAATTAACACAGA 936164
QY 565 TGGCGAGCATCAACCTGCGCGGACGAGATTCGCGCATGCTTCTGTGAAA 624
DB 936163 AGGACAAAGATTCAACTTCTGCGAGTATTCGTGAATGTCATTTCTGTGAAA 936104
QY 625 CCCAGCAAAATTCAGTCCCGTGTCTGAATGCTGTAAGAAATGACCGGCGGTGACTGA 684
DB 936103 TCCAGGAAATTCGCAACCGCGCAAGATTTGTAAGAAATGACCGGACGAGAGTTAA 936044
QY 685 ACCGACGACCGGATGATGAGTGAAGTCCGCGTATTCGTAACATTTGATATCTAC 744
DB 936043 ACCTCAGATCGTACCGTGAATGTCACAAATTCAGCATATCAGAAACATTTTGAAGATCA 935984
QY 745 GCCGATACCGCGGAATATCGCACATTCAGCGTGAAGTTATCGCTTCTGCGGTGA 804
DB 935983 TCCCATATCTCCAAATATATATATATGACTATACATGAGAAAGCTATCGTTTGTGAGAA 935924
QY 805 TCTGGAAGA 813
DB 935923 TATTGAATA 935915

```

```

RESULT 6
US-09-252-991A-16417
; Sequence 16417, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16417
; LENGTH: 837
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16417

```

```

Query Match      13.6%; Score 126; DB 4; Length 837;
Best Local Similarity 52.9%; Pred. No. 1,1e-34;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

```

```

QY 243 TCAACCTGATGATCATGATATCAATCTGCGGGTAAAGACGGTCTTCTGTAGCCGCTG 302
DB 263 TCGAGCTGTGCTGTGCTGATCCGCTGCGGCAAGACGGCTGACCTGACCCGCG 322
QY 303 AACTGCGCGACGACGCGATGTTGCGTTGATGTTCTGACTGCGCCGTCAGCAAGAGTGC 362
DB 323 AGCTGGGGGTGGCTCCGAGGTGGGATCATCTGATCACCGGGCCGACAGAGATGC 382
QY 363 ATAAATTTCTGGCTCGAAATGCTGACAGATACATACACAAACCGTTCAACCCGC 422
DB 383 ATGCATCTGCTGGCTGAGTGGCGGCGCGACGATTACGATCAAGCCGTCGACCCGC 442
QY 423 GTGACCTGACGATTCGTGACGCAACCTACTGTCCTGTCACATGATCTGGGTACTGTA 482
DB 443 GCGAATCTGTGTGCGGGGC--GAAATCTGATCCGCGGGTGCCTATGCCAGGCCA 499
QY 483 GCGAAGACGTCTGACGCTTGAAGCTACAGTTCAATGTTGGAATGGAATCATCAACA 542
DB 500 GCGCGCGCGCGCGCGGAGGCGCTCAGAGCATTCGCGCATGCTGCTGTGAGCGCCGAC 559
QY 543 GCGCTTCTGTGATCGGCGCTGATGCGAGCATCAAGCTCGCGCGAGGAGTTCCGCG 602
DB 560 GGGCGCGCTGATGACACGCGCGGACGAAACCTCTGCTCACTGCGAATTCAT 619
QY 603 CCATGCTTCACTTCTGTGAAGAACCCAGGCAAAATTGATGTCCTGTGAACTGCTGA 662
DB 620 TGTGCGCGCTTCTGCGCGACAGCGGCGATACCTTGAGCGCGACAGTGTATGAGCC 679
QY 663 AAATGACGCGCGCTGAGCTGAACCGACGACCGTACTGTAGACGATCGCGCTGA 722
DB 680 AGATCGCAACCGCGAGTGGCTGCCAGGATGCTCATCGACGTGCTGTGCGCGCC 739
QY 723 TTGCTAAATTTCAATCTACGCGCGGATACGCGGAAATCATGCGCACCATTCACG 782
DB 740 TGGTTCGAAGCTGCGCGACGACCCGCGCAACCGCACTGATCATCATCCACGCGG 799
QY 783 AAGTTATCGCTTCTGCG 800
DB 800 CCGGCTACCTGTTCAACG 817

```

```

RESULT 7
US-09-252-991A-16171/c
; Sequence 16171, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142

```

```

; SEQ ID NO 16171
; LENGTH: 2430
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16171

```

```

Query Match      13.6%; Score 126; DB 4; Length 2430;
Best Local Similarity 52.9%; Pred. No. 2,4e-34;
Matches 295; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

```

```

QY 243 TCAACCTGATGATCATGATATCAATCTGCGGGTAAAGACGGTCTTCTGTAGCCGCTG 302
DB 1406 TCGAGCTGTGCTGTGCTGATCCGCTGCGGCAAGACGGCTGACCTGACCCGCGG 1347
QY 303 AACTGCGCGACGACGCGATGTTGCGTTGATGTTCTGACTGCGCCGTCAGCAAGAGTGC 362
DB 1346 AGCTGGGGGTGGCTCCGAGGTGGGATCATCTGATCAACCGGGCGCAAGAGATGC 1287
QY 363 ATAAATTTCTGGCTCGAAATGCTGACAGATACATACACAAACCGTTCAACCCGC 422
DB 1286 ATGCATCTGCTGGCTGAGTGGCGCGCGACGATTACGTATCAAGCCGTCGAACCCGC 1227
QY 423 GTGACCTGACGATTCGTGACGCAACCTACTGTCCTGTCACATGATCTGGGTACTGTA 482
DB 1226 GCGAATCTGTGTGCGGGGC--GAAATCTGATCCGCGGGTGGCGCATGCCAGGCCA 1170
QY 483 GCGAAGACGTCTGATGCTGTAAGCTTGAAGCTTCAATGTTGGAATGGAATCATCAACA 542
DB 1169 GCGCGCGCGCGCGCGGAGGCGCTTCAAGGAGTTGGGAGTGGCTGTGAGCGCGAC 1110
QY 543 GCGCTTCTGTGATCGGCGCTGATGCGAGCATCAAGCTGCGCGAGCGAGTTCCGCG 602
DB 1109 GCGCGCGCTGATTCACACAGCGGCAACGAACCTCTGCTACACCATGCGCAATTCAT 1050
QY 603 CCATGCTTCACTTCTGTGAAGAACCCAGGCAAAATTGATGTCCTGTGAACTGCTGA 662
DB 1049 TGTGCGCGCTTCTGCGCGAACAGCGGCGATACCTGAGCGCGACGATGATGAGCC 990
QY 663 AAATGACGCGCGCTGAGCTGAACCGACGACCGTACTGTAGACGATTCGCGCTGA 722
DB 989 AGATCGCAACCGCGAGTGGCTGCCAGCGGATGCTCATGACGTGCTGTGCGCGCC 930
QY 723 TTGCTAAATTTCAATCTACGCGCGGATACGCGGAAATCATGCGCACCATTCACG 782
DB 929 TGGTTCGAAGCTGCGGACGACCGCGCAACCGCACTGATCATCATCCACGCGG 870
QY 783 AAGTTATCGCTTCTGCG 800
DB 869 CCGGCTACCTGTTCAACG 852

```

```

RESULT 8
US-09-252-991A-16309
; Sequence 16309, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 16309
; LENGTH: 3675
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-16309

```